OM of: US-08-482-402A-3_COPY_1_848 to: GenEmbl:*   out_format : pfs
Date: Nov 26, 2001 11:24 AM
About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US08482402/runat_26112001_091652_29053/app_query.fasta_1.931
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
-MNLEN=0 -MAXLEN=2000000000 -USER=US0842402_@CGN1_1_7160
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-08-482-402A-3_COPY_1_848
Query length: 848
Database: GenEmbl:*
Database sequences: 1472140
Database length: -34134837
Search time (sec): 1609.140000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	!	Documentation
gb_pr:HUMTPOC	+	4569.00	4424.87	3.4e-238	3060	!	M17755 Homo sapiens thyroid pe
gb_pat:AR092419	+	4528.00	4385.16	5.6e-236	3048	!	AR092419 Sequence 2 from paten
gb_pr:HUMTPOA	+	4528.00	4385.16	5.6e-236	3048	!	J02969 Human thyroid peroxidas
gb_pat:A92132	+	4525.00	4382.76	7.6e-236	2847	!	A92132 Sequence 1 from Patent
gb_pr:HSTPO	+	4525.00	4382.30	8.1e-236	3027	!	Y00406 Human mRNA for thyroper
gb_pr:E23825	+	4495.00	4354.51	2.8e-234	2546	!	E23825 Antigen for immunoassay
gb_pr:HSASTPO2	+	4215.00	4082.79	3.9e-219	2845	!	X17358 Human mRNA from alterna
gb_pr:HUMTPOB	+	4206.50	4073.99	1.2e-218	2877	!	J02970 Human thyroid peroxidas
gb_pr:RNTPO	+	3563.00	3449.43	7.4e-184	3237	!	X17396 Rat mRNA for thyroid pe
gb_ro:MMTPER	+	3520.00	3407.64	1.6e-181	3291	!	X60703 Mus musculus mRNA for t
gb_ro:RATTPOFR	+	3082.00	2984.38	5.9e-158	2777	!	M31655 Rat thyroid peroxidase
gb_in:AB022196	+	1927.50	1864.44	1.4e-95	3179	!	AB022197 Halocynthia roretzi m
gb_in:AB022196	+	1879.00	1818.07	5.4e-93	2915	!	AB022196 Ciona intestinalis mR
gb_ov:AF349034	+	1670.50	1613.94	1.3e-81	3846	!	AF349034 Danio rerio myelopero
gb_ov:AF378824	+	1619.00	1356.34	5.7e-79	2814	!	AF378824 Danio rerio myelond-s
gb_pat:A08802	+	1576.50	1526.77	9.1e-77	2261	!	A08802 H.sapiens gene for myel
gb_pr:HSMPOR	+	1576.50	1524.17	1.3e-76	3213	!	X04876 Human mRNA for myeloper
gb_pr:S56200	+	1575.50	1525.87	1.0e-76	2238	!	S56200 myeloperoxidase [human,
gb_pr:HUMMYP	+	1575.50	1524.85	1.2e-76	2568	!	J02694 Human myeloperoxidase m
gb_pat:E12629	+	1575.50	1523.19	1.4e-76	3215	!	E12629 cDNA encoding human nov
gb_pr:HUMMPO	+	1575.50	1523.19	1.4e-76	3215	!	M19507 Human myeloperoxidase m
gb_ro:MUSEPP	+	1564.50	1513.84	4.8e-76	2694	!	D78353 Mouse mRNA for eosinoph
gb_ro:MUSEPO	+	1559.50	1509.07	8.0e-76	2666	!	L77979 Mus musculus (clone pBM
gb_ro:MMMPO	+	1540.00	1489.97	1.0e-74	2740	!	X15313 Mouse MPO mRNA for myel
gb_in:AB028841	+	1532.00	1480.51	3.4e-74	3449	!	AB028841 Branchiostoma belcher
gb_pr:HSEP	+	1503.00	1454.62	9.6e-73	2558	!	X14346 Human mRNA for eosinoph
gb_in:AE003475	-	1438.50	1356.83	2.7e-67	300829	!	AE003475 Drosophila melanoga
gb_ov:XLU68724	+	1437.50	1391.27	3.2e-69	2512	!	U68724 Xenopus laevis polysoma
gb_htg:AC019531	+	1424.50	1353.04	4.3e-67	80254	!	AC019531 Drosophila melanogas
gb_pr:D86983	+	1421.00	1369.46	5.3e-68	5510	!	D86983 Human mRNA for KIAA0230
gb_pr:AF200348	+	1421.00	1367.86	6.5e-68	6847	!	AF200348 Homo sapiens melanoma
gb_pr:HSU39573	+	1409.50	1363.41	1.1e-67	2768	!	U39573 Human salivary peroxida
gb_om:BOVLPO	+	1404.00	1358.24	2.2e-67	2710	!	M58150 Bovine lactoperoxidase
gb_in:DM011052	+	1391.50	1341.79	1.8e-66	4871	!	U11052 Drosophila melanogaster
gb_om:AF027970	+	1382.00	1336.73	3.5e-66	2780	!	AF027970 Ovis aries airway lac
gb_om:CDR131675	+	1379.50	1334.70	4.6e-66	2634	!	AJ131675 Camelus dromedarius l
gb_pr:AC009471	+	1137.50	1069.22	2.8e-51	172445	!	AC009471 Homo sapiens BAC cl
gb_pr:HUMTPOD	+	1019.00	992.14	5.5e-47	1046	!	M55702 Human thyroperoxidase m
gb_om:SSTPOR1	-	1007.00	979.86	2.7e-46	1142	!	X04645 Porcine mRNA for thyroi
gb_htg:AF186999	-	960.50	903.11	5.0e-42	82650	!	AF186999 Homo sapiens chromos

gb_htg:AC011128   +   945.50   886.74   4.1e-41   105849   ! AC011128 Homo sapiens chr
gb_ov:AF364820   +   939.50   912.68   1.5e-42   1448   ! AF364820 Xenopus laevis per
gb_in:CELZK994   +   930.50   882.72   6.8e-41   25547   ! AF022977 Caenorhabditis el
gb_htg:AC060811   +   926.50   865.24   6.4e-40   160674   ! AC060811 Homo sapiens chr
gb_pr:HUMTPO07   +   918.50   889.95   2.7e-41   1996   ! M25707 Human thyroid peroxi

seq_name: gb_pr:HUMTPOC

seq_documentation_block:
LOCUS   HUMTPOC   3060 bp   mRNA   PRI   26-APR-1999
DEFINITION Homo sapiens thyroid peroxidase (TPO) mRNA, complete cds.
ACCESSION   M17755
VERSION   M17755.2 GI:4680720
KEYWORDS   human.
SOURCE   Homo sapiens
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 732 to 1573)
  AUTHORS   Seto,P., Hirayu,H., Magnusson,R.P., Gestautas,J., Portmann,L.,
            DeGroot,L.J. and Rapoport,B.
  TITLE   Isolation of a complementary DNA clone for thyroid microsomal
            antigen. Homology with the gene for thyroid peroxidase
  JOURNAL   J. Clin. Invest. 80 (4), 1205-1208 (1987)
  MEDLINE   88003367
REFERENCE   2 (bases 1 to 3060)
  AUTHORS   Magnusson,R.P., Chazenbalk,G.D., Gestautas,J., Seto,P., Filetti,S.,
            DeGroot,L.J. and Rapoport,B.
  TITLE   Molecular cloning of the complementary deoxyribonucleic acid for
            human thyroid peroxidase
  JOURNAL   Mol. Endocrinol. 1 (11), 856-861 (1987)
  MEDLINE   91042572
REFERENCE   3 (bases 732 to 1573)
  AUTHORS   Rapoport,B.
  TITLE   Direct Submission
  JOURNAL   Submitted (03-AUG-1993) Autoimmune Disease Unit, Cedars-Sinai
            Research Institute, 8700 Beverly Blvd, Los Angeles, CA 90048, USA
REFERENCE   4 (bases 1 to 3060)
  AUTHORS   Rapoport,B.
  TITLE   Direct Submission
  JOURNAL   Submitted (26-APR-1999) Autoimmune Disease Unit, Cedars-Sinai
            Research Institute, 8700 Beverly Blvd, Los Angeles, CA 90048, USA
REMARK   Sequence updated by submitter
COMMENT   On Apr 26, 1999 this sequence version replaced gi:339872.
FEATURES   Location/Qualifiers
     source   1..3060
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2p24"
                     /clone="PM-5"
                     /tissue_type="thyroid"
                     /note="isolated from patient with Graves' disease"
     gene   1..3060
                     /gene="TPO"
     CDS   85..2886
                     /gene="TPO"
                     /function="primary enzyme for thyroid hormone synthesis"
                     /note="similar to Homo sapiens thyroid peroxidases in
                     GenBank Accession Numbers J02969 and Y00406;
                     thyroperoxidase; membrane-associated cell surface protein;
                     primary autoantigen in human autoimmune thyroiditis
                     (Hashimoto's thyroiditis); thyroid microsomal antigen"
                     /codon_start=1
                     /product="thyroid peroxidase"
                     /protein_id="AAA61217.2"
                     /db_xref="GI:4680721"
                     /translation="MRALAVLSVTLVMACTEAFFPFISRGKELLWGKPESRVSSVLE
                     ESKRLVDTAMYATMQRNLKKRGILSGAQLLSFSKLPEPTSGVIARAAEIMETSIQAMK
                     RKVNLKTQQSOHPTDALSEDLLSIIANMSGCLPYMLPPKCPNTCLANKYRPITGACNN
                     RDHPRWGASNTALARWLPPVYEDGFSQPRGWNPGFLYNGFPLPPVREVTRHVIQVSNE
                     VVTDDRYSDLLMAWGQYIDHDIIAFTPQSTSKAAFGGGSDCQMTCENQNPCFPIQLPE
                     EARPAAGTACLPFYRSSAACGTGDQGALFGNLSTANPRQQMNGLTSFLDASTVYGSSP

```
                ALERQLRNWTSAEGLLRVHGRLRDSGRAYLPFVPPRAPAACAPEPGNPGETRGPCFLA
                GDGRASEVPSLFALHTLWLREHNRLAAALKALNAHWSADAVYQEARKVVGALHQIITL
                RDYIPRIIGPEAFQQYVGPYEGYDSTANPTVSNVFSTRAAFRFGHATIHPLVRRLDASF
                QEHPDLPGLWLHQAFFSPWTLRGGGLDPLIRGLLARPAKLQVDQLMNEELTERLFV
                LSNSSTLDLASINLQRGRDHGLPGYNEWREFCGLPRLETPADLSTAIASRSVADKILD
                LYKHPDNIDVWLGGLAENFLPRARTGPLFACLIGKQMKALRDGDWFWWENSHVFTDAQ
                RRELEKHSLSRVICDNTGLTRVPMDAFQVGKFPEDFSCSDSITGMNLEAWRETFPQDD
                KCGFPESVENGDFVHCEESGRRVLVYSCRHGYELQGREQLTCTQEGWDFQPPLCKDVN
                ECADGAHPPCHASARCRNTKGGCYCLLCADPYELGDDGRTCVDSGRLPRVTWISMSLAA
                LLIGGPAGLTSTVICRWTRTGTKSTLPISETGGGTHELRCGKHQAVGTSPQRAAAQDS
                EQESAGMEGRDTHRLPRAL"
    misc_feature    2626..2697
                /gene="TPO"
                /note="putative; transmembrane-region site"

BASE COUNT    677 a    930 c    887 g    566 t
ORIGIN

alignment_scores:
     Quality:  4569.00          Length:      848
     Ratio:       5.388         Gaps:          0
Percent Similarity: 100.000     Percent Identity: 100.000

alignment_block:
US-08-482-402A-3_COPY_1_848 x HUMTPOC    ..

Align seg 1/1  to: HUMTPOC  from: 1  to: 3060


   1 MetArgAlaLeuAlaValLeuSerValThrLeuValMetAlaCysThrGl  17
  85 ATGAGAGCCCTGGCTGTGCTCGTCGTGACGCTGGTTATGGCCTGCACAGA 134

  17 uAlaPhePheProIleSerArgGlyLysGluLeuLeuTrpGlyLysP     34
 135 AGCCTTCTTCCCCTTCATCTCGAGAGGGAAAGAACTCCTTTGGGGAAAGC 184

  34 roGluGluSerArgValSerSerValLeuGluSerLysArgLeuVal     50
 185 CTGAGGAGTCCGTGTCTTCAGCCGTCTTGGAGGAAAGCAAGCGCCTGGTG 234

  51 AspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArgGlyIl  67
 235 GACACCGCCATGTACGCCACGATGCAGAGAAACCTCAAGAAAAGAGGAAT 284

  67 eLeuSerGlyAlaGlnLeuLeuSerPheSerLysLeuProGluProThrS  84
 285 CCTTTCTGGAGCTCAGCTTCTGTCTTTTTCCAAACTTCCTGAGCCAACAA 334

  84 erGlyValIleAlaAlaArgAlaAlaGluIleMetGluThrSerIleGlnAla 100
 335 GCGGAGTGATTGCCCGAGCAGCAGATATGGAAACATCAATACAAGCG     384

 101 MetLysArgLysValAsnLeuLysThrGlnIleHisSerProThrAs    117
 385 ATGAAAAGAAAAGTCAACCTGAAAACTCAACAATCACAGCATCCAACGA  434

 117 pAlaLeuSerGluGluSerIleIleAlaAlaAsnMetSerGlyCysLysL 134
 435 TGCTTTATCAGAAGATCTGCTGAGCATCATTGCAAACATGTCTGGATGTC 484

 134 euProTyrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLys 150
 485 TCCCTTACATGCTGCCCCCAAAATGCCCAAACACTTGCCTGGCAACAAA  534

 151 TyrArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpGl 167
 535 TACAGGCCCATCACAGGAGCTTGCAACAACAGAGACCACCCCAGATGGGG 584

 167 yAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspG 184
 585 CGGCTTCCAACACGGCCCTGGCACGATGGCTCCCTCCAGTCATGAGGACG 634

 184 lyPheSerGlnProArgGlyTrpAsnProGlyPheLeuTyrAsnGlyPhe 200


 635 GCTTCAGTCAGCCCCGAGGCTGGAACCCCGGCTTCTTGTACAACGGGTTC 684

 201 ProLeuProProValArgGluValThrArgHisValIleGlnValSerAs 217
 685 CCACTGCCCCCGGTCCGGGAGGTGACAAGACATGTCATTCAAGTTTCAAA 734

 217 nGluValValThrAspAspAspArgTyrSerAspLeuMetAlaTrpG    234
 735 TGAGGTTGTCACAGATGATGACCGCTATTCTGACCTCCTGATGGCATGGG 784

 234 lyGlnTyrIleAspHisAlaPheThrProGlnSerThrSerLys       250
 785 GACAATACATCGACCACGCATCCGTTCACACCACAGAGCACCAGCAAA   834

 251 AlaAlaPheGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAs 267
 835 GCTGCCTTCGGGGAGGGGTCGACATGACTTGTGAGAACCAAAA        884

 267 nProCysPheProIleLeuProGluGlnAlaArgProAlaAlaAlaGlyT 284
 885 CCCATGTTTTCCCATACAACTCCCGAGGAGCCCGGCCCGGCGGGCA     934

 284 hrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyThrGlyAsp 300
 935 CCGCCTGCTGCCCTTCTACCGCTCTTCCGGCCGCCTGCGGCACCGGGGAC 984

 301 GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe 317
 985 CAAGGCGGCTCTTTGGAACCTCTCCACGGCCAACCCGAGGCAGCAGAT  1034

 317 tAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGlySerSerP 334
1035 GAACGGGTTGACCTCGTTCCTGGACGCGTCCACCGTATGGCAGCTCCC  1084

 334 roAlaLeuGluArgGlnLeuArgAsnTrpThrSerAlaGluGlyLeuLeu 350
1085 CGGCCCTAGAGAGGCAGTTGCGAAACTGGACCAGTGCCGAAGGGCTGCTC 1134

 351 ArgValHisGlyArgLeuArgAspSerGlyArgAlaTyrLeuProPheVa 367
1135 CGGGTCCACGGCCGCCTCCGGGACTCCGGGCCGCCTACCTCCCTTCGT  1184

 367 lProProArgAlaProAlaAlaCysAlaProGluProGlyAsnProGlyG 384
1185 GCCGCCACGCGCGCTGCGCCTGCGCCCGAGCCCGGCAACCCCGGAG    1234

 384 luThrArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerGluVal 400
1235 AGACCCGAGGCCGGGCCCTGCTTCCTGGCCGGAGACGGCCGCCAGCGAGGTC 1284

 401 ProSerLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLe 417
1285 CCCTCCCTGACGGCACTGCACACCCTGTGGCTGCGGGAGCACAACCGCCT 1334

 417 uAlaAlaAlaLeuLysAlaLeuAsnAlaHisTrpSerAlaAspAlaValT 434
1335 GGCCGCGGCGCTCAAGGCCCTCAATGCGCACTGGAGCGCCGACGCGGTGT 1384

 434 yrGlnGluAlaArgLysValValGlyAlaLeuHisGlnIleIleThrLeu 450
1385 ACCAGGAGGCGCAAGGTCGTGGGCGCTCTGCACCAGATCATCACCCTG  1434

 451 ArgAspTyrIleProArgIleLeuGlyProGluAlaPheGlnGlnTyrVa 467
1435 AGGGATTACATCCCCAGGATCCTGGGACCCGAGGCCTTCCAGCAGTACGT 1484

 467 lGlyProTyrGluGlyTyrAspSerThrAlaAsnProThrValSerAsnV 484
1485 GGGTCCCTATGAAGGCTATGACTCACCCGCCAACCCCACTGTGTCAACG  1534

 484 alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu 500
```

801 AlaAspGlyAlaHisProProCysHisAlaSerAlaAlaArgCysArgAsnTh 817
    |||||||||||||||||||||||||||||||||||||||||||||||||||||
2485 GCAGACGGTGCCCACCCCCCTGCCACGCCCTGCGAGTGCGAGAAACAC 2534

817 rLysGlyGlyPheGlnCysLeuCysAlaAspProTyrGluGlyLysAspA 834
    |||||||||||||||||||||||||||||||||||||||||||||||||||
2535 CAAAGGGCGGCTTCCAGTGTCTCTGCGCCGGACCCCTACGAGTAGGAGACG 2584

834 spGlyArgThrCysValAspSerGlyArgLeuProArgValThr 848
    ||||||||||||||||||||||||||||||||||||||||||||
2585 ATGGGAGAACCTGCGTAGACTCCGGGAGGCTCCCGGGGTGACT 2628

seq_name: gb_pat:AR092419

seq_documentation_block:
LOCUS   AR092419   3048 bp   DNA   PAT   08-SEP-2000
DEFINITION   Sequence 2 from patent US 5998153.
ACCESSION   AR092419
VERSION   AR092419.1   GI:10019173
KEYWORDS   .
SOURCE   Unknown.
  ORGANISM   Unknown.
             Unclassified.
REFERENCE   1 (bases 1 to 3048)
  AUTHORS   Baker,J.R. Jr. and Koenig,R.J.
  TITLE   Thyroid peroxidase epitopic regions
  JOURNAL   Patent: US 5998153-A 2 07-DEC-1999;
FEATURES   Location/Qualifiers
     source   1..3048
                     /organism="unknown"
BASE COUNT   671 a   930 c   881 g   566 t
ORIGIN

alignment_scores:
     Quality: 4528.00   Length: 848
     Ratio: 5.365   Gaps: 0
Percent Similarity: 99.528   Percent Identity: 99.175

alignment_block:
US-08-482-402A-3_copy_1_848 x AR092419

Align seg 1/1   to: AR092419   from: 1   to: 3048

   1 MetArgAlaLeuAlaValLeuSerValThrLeuValMetAlaCysThrGl 17
     |||||||||||||||||||||||||||||||||||||||||||||||||
  73 ATGAGAGCGCTCGCTGTGCTCTCTGTCACGCTGGTTATGGCCTGCACAGA 122

  17 uAlaPhePheProPheLeuSerArgGlyLysGluLeuTrpGlyLysP 34
     |||||||||||||||||||||||||||||||||||||||||||||||
 123 AGCCTTCTTCCCCTTTCATCTCGAGAGGGAAAGAACTCCTTTGGGGAAAGC 172

  34 roGluGluSerArgValSerSerValLeuGluSerLysArgLeuVal 50
     ||||||||||||||||||||||||||||||||||||||||||||||
 173 CTGGAGGAGTCTCGTGTCTCTAGCGTCTTGGAGAAGAAGCAGCCTGGTG 222

  51 AspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArgGlyIl 67
     |||||||||||||||||||||||||||||||||||||||||||||||||
 223 GACACCGCCATGTACGCCACGATGCAGAGAAACCTCAAGAAAAGAGGAAT 272

  67 eLeuSerGlyAlaGlnLeuLeuSerPheSerLysLeuProGluProThrS 84
     |||||||||||||||||||||||||||||||||||||||||||||||||
 273 CCTTTCTCCAGCTCAGCTTCTGTTTTTCCAAACTTCCTGAGCCAACAA 322

  84 erGlyValIleAlaArgAlaAlaGluIleMetGluThrSerIleGlnAla 100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 323 GCGGAGTGATTGCCCGAGCCGCAGAGATAATGGAAACATCAATACAAGCG 372

 101 MetLysArgLysValAsnLeuLysThrGlnHisSerGlnHisProThrAs 117
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 373 ATGAAAAGAAAGTCAACCTGAAAACTCAACAATCACAGCATCCAACGGA 422

1535 TGTTCTCCACAGCCGCCTTCCGCTTCGGCCATGCCACGATCCACCCGCTG 1584

 501 ValArgArgAspAlaSerPheGlnGluHisProLeuProGlyLeu 517
     ||||||||||||||||||||||||||||||||||||||||||||
1585 GTGAGGAGGCTGGACGCCAGCTTCCAGGAGCACCCGACCTGCCCGGCT 1634

 517 uTrpLeuHisGlnAlaPhePheSerProTrpThrLeuArgArgGlyG 534
     ||||||||||||||||||||||||||||||||||||||||||||||
1635 GTGGCTGCCACCAGGCTTTCTTCAGCCACATTACTCCGTGGAGGTG 1684

 534 lyLeuAspProLeuIleArgGlyLeuLeuAlaArgProAlaLysLeuGln 550
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1685 GTTTGGACCCACTAATACGAGGCCTTCTTGCAAGACCAGCCAAACTGCAG 1734

 551 ValGlnAsnGlnLeuMetCysAsnGluLeuThrGluLeuThrGluValPheValLe 567
     |||||||||||||||||||||||||||||||||||||||||||||||||||||
1735 GTGCAGGATGCTGATGAACGAGGAGGCTGACGGAAAGGCTCTTTGTGCT 1784

 567 uSerAsnSerSerThrLeuAspLeuAlaSerIleAsnLeuGlnArgGlyA 584
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1785 GTCCAATTCCAGACCCTTGGATCTGGCGTCCATCAACCTGCAGAGGGGCC 1834

 584 rgAspHisGlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeu 600
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1835 GGGACCACGGGCTGCCAGGTTACAATGAGTGGAGGGAGTTCTGCGGCCTG 1884

 601 ProArgLeuGluThrProAlaAspAspLeuSerThrAlaIleAlaSerArgSe 617
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1885 CCTCGCCTGGAGACCCCCGCTGACGATCTGAGCACAGCCATCGCCAGCAGGAG 1934

 617 rValAlaAlaSerLysHisIleLeuAspLeuTyrLysHisProAspAsnIleAspV 634
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1935 CGTGGCCGCAAGATCCTGGACTTGTACAAGCATCCTGACAACATCGATG 1984

 634 alTrpLeuGlyGlyLeuGluAlaGluAsnPheLeuProArgAlaArgThrGly 650
     |||||||||||||||||||||||||||||||||||||||||||||||||
1985 TCTGGCTGGGAGGCTTAGCTGAAAACTTCCTCCCCAGGGCTCGGACCAGGG 2034

 651 ProLeuPheAlaCysLeuLeuGlyLysGlnMetLysAlaLeuArgAspGl 667
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2035 CCCCTGTTTGCCTGTCTCATTGGGAAGCAGATGAAGGCTCTGCGGGACGG 2084

 667 yAspTrpPheThrPheTrpGluAsnSerHisValPheThrAspAlaGlnArgA 684
     |||||||||||||||||||||||||||||||||||||||||||||||||
2085 TGACTGGTTTGGTTTTGGGAGAACAGCCACGTCTTCACGGATGCACAGAGGC 2134

 684 rgGluLeuLysHisSerLeuSerArgValIleCysAspAsnThrGly 700
     ||||||||||||||||||||||||||||||||||||||||||||||
2135 GTGAGCTGAAGCACTCCCTGTCTCGGGTCATCTGTGACAACACTGGC 2184

 701 LeuThrArgValProMetAspAlaPheGlnAsnValGlyLysPheProGluAs 717
     |||||||||||||||||||||||||||||||||||||||||||||||||
2185 CTCACCAGGGTGCCCATGGATGCCTTCCAAGTGGGCAAATTCCCCGAGA 2234

 717 pPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgG 734
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2235 CTTTGAGTCTTGTGACAGCATCACTGGCATGAACCTGGAGGCCTGGAGGG 2284

 734 luThrPheGlnGlnAspAspLysCysGlyPheProGluSerValGluAsn 750
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2285 AAACCTTTCCTCAAGACCAACAAGTGTGGCTTCCCAGAGAGCGTGGAGAAT 2334

 751 GlyAspPheValHisCysCysGlyGluSerGlyArgArgValLeuValTyrSe 767
     ||||||||||||||||||||||||||||||||||||||||||||||||
2335 GGGGACTTTGTGCACTGTGCACTGTGGGGAGTCTGGGAGGCGCGTGCTGGTGTATTC 2384

 767 rCysArgHisGlyTyrGluLeuGlnGlyArgGluGlnLeuThrCysThrG 784
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2385 CTGCAGGCAGGGTATGAGCTGCAAGGCCGGGAGCAGCTCACTTGCACCC 2434

 784 lnGluGlyTrpAspProPheProLeuCysLysAspValAsnGluCys 800
     ||||||||||||||||||||||||||||||||||||||||||||||
2435 AGGAAGGGTGGGGATTTCCAGCCTCCCTGTGCAAAGATGTGAACGAGTGT 2484

```
117 pAlaLeuSerGluAspLeuLeuSerIleIleIleAlaAsnMetSerGlyCysL 134      1323 ||||||||||||||||||||||||||||||||||||||||||||||||||||| pAlaLeuSerGluAspLeuLeuSerIleIleIleAlaAsnMetSerGlyCysL 1372
423 TGCTTTATCAGAGATCTCTGAGCATCATTGCAAACATGTCTGGATGTC 472             GGCCGCGGCTCAAGGCCTCAAGGCCCTCAATGCCACTGCACTGGAGCGCGGACGCGCGTGT

134 euProTyrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLys 150          1373 ||||||||||||||||||||||||||||||||||||||||||||||||||||| yrGlnGluAlaArgLysValValGlyAlaAlaLeuHisGlnIleIleThrLeu 1422
473 TCCCTTACACGTCTGCCCCAAAATGCCCAAACACTTGCCTGGCGAACAAA 522           ACCAGGAGGCGCCCCAAGGTCGTGGCGGCCGCTCTGCACCAGATCATCACCCTG

151 TyrArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpGl 167          1423 ||||||||||||||||||||||||||||||||||||||||||||||||||||| ArgAspTyrIleProArgIleLeuGlyProGluAlaPheGlnInTyrVa 1467
523 TACAGGCCCCATCAACAGGAGGCTTGCAACACAGAGACCCACCCCAGATGGGG 572        AGGGATTACATCCCCAGGATCCTGGGACCCGGAGGCCTTCCAGCAGTACGT

167 1GlyProTyrGluGlyTyrAspSerThrAlaAsnProThrValSerAsnV 184         1473 ||||||||||||||||||||||||||||||||||||||||||||||||||||| 1GlyProTyrGluGlyTyrAspSerThrAlaAsnProThrValSerAsnV 1522
573 CGCTCCCAACACCGCCCTGCACGATGGCTCCCCAGTCTATGAGGACG 622              GGGTCCCTATGAGGGCTATGACTCCACCGCCAACCCACTGTCTCCAACG

184 alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu 200         1523 ||||||||||||||||||||||||||||||||||||||||||||||||||||| alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu 1572
623 GCTTCAGTCAGCCCCGGAGGCTGGAACCCCGGCTTCTTGACAACGGGTTC 672          TGTTCTCCACAGCCGCCTTCCGCTTCGGCCATGCCACGATCCACCCGCTG

201 ProLeuProProValArgGluValThrArgHisValIleGluGlnValSerAs 217      1573 ||||||||||||||||||||||||||||||||||||||||||||||||||||| ValArgLeuAspSerPheGlnGluHisProAspLeuProGlyLe 517
673 CCACTGCCCCCGGTCCGGGAGGTGACAAGACATGTCATTCAAGTTCAAA 722          GTGAGGAGGCTGGACGCCAGCTTCCAGGAGCACCCGGACCTGCCCGGGCT

217 nGluValValThrAspAspAspAspArgTyrSerAspLeuLeuMetAlaTrpG 234     1623 ||||||||||||||||||||||||||||||||||||||||||||||||||||| uTrpLeuHisGlnAlaPhePheSerProThrTrpLeuLeuArgGlyGlyG 534
723 TGAGGTTGCACAGATGATGACCGTATTCTGACCTCCTATGGCATGGG 772            GTGGTGCACCAGCTTTCTTCAGCCCATGGACATTACTCCGTGGAGGTG

234 lyGlnTyrIleAspHisAspAlaPheThrProGlnSerThrSerLys 250          1673 ||||||||||||||||||||||||||||||||||||||||||||||||||||| lyLeuAspProLeuIleArgGlyLeuLeuAlaArgProAlaLysLeuGln 550
773 GACAATACATCGACCACGACATCGCGTTCACACCACAGAGACCAGCAAA 822          GTTTGGACCCACCACTAATACGAGGCCTTCTGCAAGCACCAGCCAAACTGCAG

251 AlaAlaPheGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAs 267        1723 ||||||||||||||||||||||||||||||||||||||||||||||||||||| ValGlnAspGlnLeuMetLeuMetGluAsnGluGluArgLeuPheValIe 567
823 GCTGCCTTCGGGGAGGGGTCTGACTGGGCCCAGATGACTTGTGAGAACCAAAA 872      GTGCAGGATCAGCTGATGAACGAGGAGCTTGACGGAAAGGCTCTTTGTGCT

267 nProCysPheProIleGlnLeuProGluGlnAlaArgProAlaAlaGlyT 284        1773 ||||||||||||||||||||||||||||||||||||||||||||||||||||| uSerAsnSerSerThrLeuAspLeuAlaSerIleAsnLeuGlnArgGlyA 584
873 CCCATGTTTTCCCATACAACTCCCCGGAGGAGCCCGGCCGCCGCGGGGCA 922         GTCCAATTCCAGCACCTTGGATCTGGCGTCCATCAACCTGCAGAGGGGCC

284 hrAlaCysLeuProTheTyrArgSerSerAlaAlaCysGlyThrGlyAsp 300        1823 ||||||||||||||||||||||||||||||||||||||||||||||||||||| rgAspHisGlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeu 600
923 CCGGCCTGTCTGCCCTTCTACCGGTCTTCGGCGGCACCGGGGAC 972              GGGACCACGGGCTGCCAGGTTACAATGAGTGAGGAGTTCTGCGGCCTG

301 GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe 317        1873 ||||||||||||||||||||||||||||||||||||||||||||||||||||| ProArgLeuGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSe 617
973 CAAGGGGCGCTCTTTGGGAACCTGTCCACGGCCAACCCGCGGCAGCAGAT 1022        CCTCGCCTGGAGACCCCCGGAGACCTGAGCACAGCCATCGCCAGCAGGAG

317 tAsnGlyLeuThrSerPheLeuAspAlaCysAlaSerThrValTyrGlySerSerP 334  1923 ||||||||||||||||||||||||||||||||||||||||||||||||||||| rValAlaAspLysIleLeuLeuAspLeuTyrLysHisProAspAsnIleAspV 634
1023 GAACGGGTTGACCTCGTTCCTGGACGCGTGTGCCAGCACCGTATGGCAGCTCCC 1072    CGTGGCCGACAAGATCCTGGACTTGTACAAGCATCCTGACAACATCGATG

334 roAlaLeuGluArgAsnTrpThrSerAlaGluGluValAspLeuLeu 350         1973 ||||||||||||||||||||||||||||||||||||||||||||||||||||| alTrpLeuGlyLeuAlaGluAsnPheLeuProArgArgAlaArgThrGly 650
1073 CGGGCCACCGGGTTGACGCAGGCTGCGGAACTGGACCAGTGCCGAAGGGCTGCTC 1122   TCTGGCTGGAGGCTTAGCTGAAAACTTCCTCCCCAGGCCTCGGACAGGG

351 ArgValHisGlyArgGlyArgAspSerGlyTyrArgAlaTyrLeuProPheVa 367   2023 ||||||||||||||||||||||||||||||||||||||||||||||||||||| ProLeuPheAlaCysLeuIleGlyLeuIleGlnMetLysGlnAlaLeuArgAspGl 667
1123 CGGCTCCACGGCGCCCTGCGCCGGCCTACCTGCCCTTCGT 1172                 CCCCTGTTTGCCTGCCTGCTCATTGGGAAGCAGAGTGAAGGCCTCGCGGGATGG

367 lProProArgAlaProAlaAlaCysAlaProGlyGluProGlyAsnProGlyG 384  2073 ||||||||||||||||||||||||||||||||||||||||||||||||||||| yAspTrpPheTrpTrpGluAsnSerHisValPheThrAspAlaGlnInArgA 684
1173 GCCGCCACGGAGGGGTTGACCTCGGTGCCGTGCCTGCGCCCGCATCCCCGGAG 1222    TGACTGGTTTTGGTGGGAGGAAAACAGCCACCGTCTTCACGGATGCCACAGAGGC

384 luThrArgGlyProCysPheLeuLeuAlaGlyAlaGlyArgAlaSerGluVal 400  2123 ||||||||||||||||||||||||||||||||||||||||||||||||||||| rgGluLeuGluLysHisSerLeuSerArgValIleCysAspAsnThrGly 700
1223 AGACCCGGGGCCCGGTGCTTCCTGCGCGGGGGCGGCGCCAGCCAGGTC 1272       GTGAGCTTGGAGAAGCACTCCCCTGTCTCGGGTCATCTGTGACAACACTGGC

401 ProSerLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgIe 417   2173 ||||||||||||||||||||||||||||||||||||||||||||||||||||| LeuThrArgValProMetAspAlaPheGlnInValGlyLysPheProGluAs 717
1273 CCCTCCCTGACGGCGCTGCCACACCGCTGTGCTCCGCGAGCACAACCGCCT 1322     CTCACCAGGGTGCCCATGGATGCCTTCCAAGTCGGCAAATTCCCTGAAGA

417 uAlaAlaLeuAlaLeuLysAlaLeuAsnAlaAlaHisTrpSerArgAlaAspAlaValT 434  2223 ||||||||||||||||||||||||||||||||||||||||||||||||||||| pPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgG 734
```

```
                                          VVTDDRYSDLLMAWQOYIDHDIAFFPQSTSKAAFGGADCQMTCENQNPCFPIQLPE
                                          EARPAAGTACLPFYRSSAACGTGDQGALFGNLSTANPRQQMNGLTSFLDASTVYGSSP
                                          ALERQLRNWTSAEGLLRVHARLRDSGRAYLPFVPPRRPAACAPEPGIPGETRGPCFLA
                                          GDGRASEVPSLTALHTLWLREHNRLAAALKALNAHWSADAVVQEARKVVGALHQIITL
                                          RDYIPRILGPEAFQQYVGPYEGYDSTANPTVSNVFSTAAFRFGHATIHPLVRRIDASF
                                          QEHPDLPGLWLHQAFFSPWTLRGGGLDPLIRGLLARPAKLQVDQLMNEELTERLFV
                                          LSNSSTLDLASINLQRGRDHGLPGYNEWREFCGLPRLETPADLSTAIASRSVADKILD
                                          LYKHPDNIDVWIGGLAENFLPRARTGPLFACLIGKQMKALRDGDWFWWENSHVFTDAQ
                                          RRELEKHSLSRVICDNTGLTRVPMDAFQVGKFPEDFESCDSIPGMNLEAWRETFPQDD
                                          KCGFPESVENQDFVHCEESGRRVLVYSCRHGYELQGREQLTCTQEGWDFQPPLCKDVN
                                          ECADGAHPPCHASARCRNTKGGFQCLCADPYELGDDGRTCVDSGRLPRATWISMSLAA
                                          LLIGGFAGLTSVICWRTWTGTKSTLPISETGGGTPELRCGRKHQAVGTSPQRAAAQDS
                                          EQESAGMEGRDTHRLPRAL"

BASE COUNT     671 a    930 c    881 g    566 t
ORIGIN      Chromosome 2pter-q11.

alignment_scores:
         Quality:  4528.00              Length:   848
         Ratio:    5.365                Gaps:       0
Percent Similarity:  99.528   Percent Identity:  99.175

alignment_block:
US-08-482-402A-3_COPY_1_848 x HUMTPOA

Align seg 1/1  to: HUMTPOA  from: 1  to: 3048

    1 MetArgAlaLeuAlaValLeuSerValThrLeuValMetAlaCysThrG1  17
      |||||||||||||||||||||||||||||||||||||||||||||||||
   73 ATGAGAGCGCTCGCTGCTGTCCTGTCTGTCACGCTGGTTATGGCCTGCACAGA  122

   17 uAlaPhePheProPheLeuSerArgGlyLysGluLeuLeuTrpGlyLysP  34
      |||||||||||||||||||||||||||||||||||||||||||||||||
  123 AGCTTCTTCCCCTTCATCTCGAGAGGGAAAGAACTCCTTTGGGGAAAGC  172

   34 roGluGluSerArgValSerSerValLeuGluSerLysArgLeuVal  50
      ||||||||||||||||||||||||||||||||||||||||||||||
  173 CTGAGGAGTCTCGTGTCTCTAGCCGTCTTGGAGAGAAAGAAGCGCCTGGTG  222

   51 AspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArgGlyI1  67
      |||||||||||||||||||||||||||||||||||||||||||||||||
  223 GACACCGCCATGTACGCCACGATGCAGAGAAACCTCAAGAAAAGAGGAAT  272

   67 eLeuSerGlyAlaGlnLeuLeuSerPheSerLysLeuProGluProThrS  84
      |||||||||||||||||||||||||||||||||||||||||||||||||
  273 CCTTCTCCAGCTCAGCTTCGTCTTTTTCCAAACTTCCTGAGCCAACAA  322

   84 erGlyValIleAlaArgAlaAlaAlaGluIleMetGluThrSerIleGlnAla  100
      |||||||||||||||||||||||||||||||||||||||||||||||||
  323 GCGGAGTGATTGCCCGAGCAGCAGCAGAGATAATGGAAACATCAATACAAGCG  372

  101 MetLysArgLysValAsnLeuLysThrGlnSerGlnHisProThrAs  117
      |||||||||||||||||||||||||||||||||||||||||||||||||
  373 ATGAAAAGGAAAGTCAACCTGAAAACTCAAAACTCAACAATCACAGCATCCAACGGA  422

  117 pAlaLeuSerGluAspLeuLeuSerIleIleAlaAsnMetSerGlyCysL  134
      |||||||||||||||||||||||||||||||||||||||||||||||||
  423 TGCTTTATCAGAAGATCTGCTGAGCATCATTGCAAACATGTCTGATGTC  472

  134 euProTyrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLys  150
      |||||||||||||||||||||||||||||||||||||||||||||||||
  473 TCCCTTACATGTCGCCCCCAAATGCCCAAACACTTGCCTGGCGAACAAA  522

  151 TyrArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpG1  167
      |||||||||||||||||||||||||||||||||||||||||||||||||
  523 TACAGGCCCATCACAGGAGCTTGCAACACAGACCACCCACAGATGGGG  572

  167 yAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspG  184
      |||||||||||||||||||||||||||||||||||||||||||||||||
  573 CGCCTCCAACACGGCCCTGGCACGATGGCTCCCCTCCAGTCTATGAGGACG  622

  184 lyPheSerGlnProArgGlyTrpAsnProGlyPheLeuTyrAsnGlyPhe  200
      |||||||||||||||||||||||||||||||||||||||||||||||||
```

```
 2223 CTTTGAGTCTTGTGACAGCATCCCTGGCATGAACCTGGAGGCCTGGAGGG  2272
      |||||||||||||||||||||||||||||||||||||||||||||||||

  734 luThrPheProGlnAspLysCysGlyPheProGluSerValGluAsn  750
      |||||||||||||||||||||||||||||||||||||||||||||
 2273 AAACCTTCCTCAGACACAAGTGTGGCTTCCCAGAGACGCGTGGAGAAT  2322

  751 GlyAspPheValHisCysGluGluSerGlyArgArgValLeuValTyrSe  767
      |||||||||||||||||||||||||||||||||||||||||||||||||
 2323 GGGGACTTTGTGCACTGTGAGGAGTCTGGGAGGCGCGTGCTGGTGTATTC  2372

  767 rCysArgHisGlyTyrGluLeuGlnGlyArgGluGlnLeuThrCysThrG  784
      |||||||||||||||||||||||||||||||||||||||||||||||
 2373 CTGCCGGCACGGGTATGAGCTCCAAGGCCGGGAGCAGCTCACTTGCACCC  2422

  784 lnGluGlyTrpAspPheGlnProProLeuCysLysAspValAsnGluCys  800
      |||||||||||||||||||||||||||||||||||||||||||||
 2423 AGGAAGGATGGGATTTCCAGCCTCCCCTGTGCAAAGATGTGAACGAGTGT  2472

  801 AlaAspGlyAlaHisProProCysHisAlaSerAlaArgCysArgAsnTh  817
      |||||||||||||||||||||||||||||||||||||||||||||||||
 2473 GCAGACGGTGCCCACCCCCCCTGCCACGCCCTCTGCGAGGTGCAGAAACAC  2522

  817 rLysGlyGlyPheGlnCysLeuCysAlaAspProTyrGluLeuGlyAspA  834
      |||||||||||||||||||||||||||||||||||||||||||||||||
 2523 CAAAGGCGGCTTCCAGTGTCTCTGCGCGGACCCCTACGAGTTAGGAGACG  2572

  834 spGlyArgThrCysValAspSerGlyArgLeuProArgValThr  848
      ||||||||||||||||||||||||||||||::::||
 2573 ATGGGAGAACCTGCGTAGACTCCGGGAGCCTCCCTCGGGCGACT  2616
```

```
seq_name: gb_pr:HUMTPOA

seq_documentation_block:
LOCUS       HUMTPOA       3048 bp    mRNA          PRI       14-JAN-1995
DEFINITION  Human thyroid peroxidase mRNA, clone phTPO-2.8.
ACCESSION   J02969
VERSION     J02969.1  GI:339866
KEYWORDS    thyroid peroxidase.
SOURCE      Human thyroid gland (from patients with Grave's disease), cDNA to
            mRNA, clone phTPO-2.8.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3048)
AUTHORS     Kimura,S., Kotani,T., McBride,O.W., Umeki,K., Hirai,K., Nakayama,T.
            and Ohtaki,S.
TITLE       Human thyroid peroxidase: complete cDNA and protein sequence,
            chromosome mapping, and identification of two alternately spliced
            mRNAs
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 84 (16), 5555-5559 (1987)
MEDLINE     87289643
COMMENT     Draft entry and computer-readable sequence for [1] kindly provided
            by S.Kimura, 09-SEP-1987.
FEATURES             Location/Qualifiers
     source          1..3048
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="2pter-p24"
     mRNA            <1..3048
                     /note="TPO mRNA"
     gene            73..2874
                     /gene="TPO"
     CDS             73..2874
                     /gene="TPO"
                     /note="thyroid peroxidase"
                     /codon_start=1
                     /db_xref="GDB:G00-120-446"
                     /protein_id="AAA61215.1"
                     /db_xref="GI:339867"
                     /translation="MRALAVLSVTLVMACTEAFFPFISRGKELLWGKPEESRVSSVLE
                     ESKRLVDTAMYATMQRNLKKRGILSPAQLLSFSKLPEPTSGVIARAAEIMETSIQAMK
                     RKVNLKTQQSQHPTDALSEDLLSIIANMSGCLPYMLPPKCPNTCLANKYRPITGACNN
                     RDHPRWGASNTALARWLPPVYEDGFSQPRGWNPGFLYNGFPLPPVREVTRHVIQVSNE
```

Mon Nov 26 15:43:03 2001   us-08-482-402a-3_copy_1_848.rge

```
623 GCTTCAGTCAGCCCCGAGGCTGGAACCCCGGCTTCTTGTACAACGGGTTC 672

201 ProLeuProProValArgThrArgGluValThrIleValIleGlnValSerAs 217
    ||||||||||||||||||||||||
673 CCACTGCCCCCGGTCCGGGAGGTGACAAGACATGTCATTCAAGTTTCAAA 722

217 nGluValValThrAspAspAspArgTyrSerAspLeuMetAlaTrpG 234
    ||||||||||||
723 TGAGGTTGTCACAGATGACCCGTATTCTGACCTCCTGATGGCATGGG 772

234 lyGlnTyrIleAspHisAlaPheThrProGlnSerThrSerLys 250
    ||||||||||||
773 GACAATACATCGACCACGACATCGCGTTCACACCACAGAGCACCAGCAAA 822

251 AlaAlaPheGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAs 267
    |||||||||||||::::|||||||||||||||
823 GCTGCCTTCGGAGGCGGGAGCGACTGTCAGATGACTTGTGAGAACCAAAA 872

267 nProCysPheProIleGlnLeuProGluGluAlaArgProAlaAlaGlyT 284
    |||||||||||||||||
873 CCCATGTTTTCCATACAACTCCCGAGGAGGCCCGGCCGCGGGCA 922

284 hrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyThrGlyAsp 300
    ||||||||||||
923 CCGCCGTGTCTGCCCTTCTACCGGTCTTCGGCGCCTGCGGCACCGGGGAC 972

301 GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe 317
    ||||||||||||
973 CAAGGCGGCTCTTTGGGAACCTGTCCACGGCCAACCCGCGGCAGCAGAT 1022

317 tAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGlySerSerP 334
    ||||||||||||
1023 GAACGGGTTGACCTCGTTCCTGGACGCGTCCACCGTGTATGGCAGCTCCC 1072

334 roAlaLeuGluGluArgGlnLeuArgAsnTrpThrSerAlaGluGluVal 350
    |||||||||    |||||
1073 CGGCCCTAGAGGAGAGGCAGCTGCGGAACTGGACCAGTGCCGAAGGGGTC 1122

351 ArgValHisGlyArgLeuArgAspAspSerGlyArgAlaTyrLeuProPheVa 367
    ||||||||::::||||||
1123 CGCGTCCACGGGCGGCTCCGGGACTCCGGCCGCCTACCTGCCCTTCGT 1172

367 lProProArgAlaAlaProAlaAlaCysProAlaProGluProGlyAsnProGlyG 384
    |||||||     |||||
1173 GCCCCCACGGCGGCCGCCTGCCGGCCCGGAGCCCGGCATCCCGGAG 1222

384 luThrArgGlyArgGlyProCysPheLeuAlaGlyAlaGlyAspAlaSerGluVal 400
    ||||||||||||
1223 AGACCCCGCGGGCGCGGGCCCTGCTTCCTGGCCGGAGCGGGGCGCCAGCGAGGTC 1272

401 ProSerLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLe 417
    ||||||||::::|||||
1273 CCCTCCCTGACGGCACTGCACACCCTGTGGCTGCGCGAGCACAACCGCCT 1322

417 uAlaAlaAlaLeuLysAlaLeuAsnAlaAlaHisTrpSerAlaAlaAspAlaValT 434
    ||||||||||||
1323 GGCCGCGGCGCTCAAGGCCCTCAATGCCGCCCACTGGAGCGCGGACGCCGTGT 1372

434 yrGlnGluAlaArgLysValValGlyAlaAlaLeuHisGlnIleIleThrLeu 450
    ||||||||||||
1373 ACCAGGAGGCGCGCAAGGTCGTGGGCGCCGCTCTGCACCAGATCATCACCCTG 1422

451 ArgAspTyrIleProGluArgIleLeuProGluAlaPheGlnInTyrArgVa 467
    ||||||||||||
1423 AGGGATTACATCCCCGAGCGCATCCTGGACCCGAGGCCTTCCAGCAGTACGT 1472

467 lGlyProTyrGluGlyTyrAspProSerThrAlaAsnProThrValSerAsnV 484
    ||||||||||||
1473 GGGTCCCATATGAGGGCTATGACCCATCCACCGCCAACCCCACTGTCTCCAACG 1522

484 alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu 500
    ||||||||||||
1523 TGTTCTCCACAGCGGCCTTCCGGTTCGGCCATGCCACGATCCACCCGCTG 1572
```

```
501 ValArgArgLeuAspAlaSerPheGlnGluHisProAspLeuProGlyLe 517
    ||||||||||||||||||||||||||||||||||||||||||||||
1573 GTGAGGAGGCTGGACGCCAGCTTCCAGGAGCACCCGACCTGCCCGGGCT 1622

517 uTrpLeuHisGlnAlaPheSerProTrpThrLeuLeuArgGlyGlyG 534
    ||||||||||||||||||||||||||||||||||||||||
1623 GTGGCTGCACCAGCTTCTTCAGCCACTGGACATTACTCCGTGGAGGTG 1672

534 lyLeuAspProLeuIleArgGlyLeuAlaArgProAlaLysLeuGln 550
    |||||||||||||||||||||||||||||||||||||
1673 GTTTGGACCCACTAATACGAGGCCTTCTTGCAAGACCAGCCAAACTGCAG 1722

551 ValGlnAspGlnLeuMetAsnGluGluLeuThrGlyArgLeuPheValLe 567
    ||||||||||||||||||||||||||||||||||||||||
1723 GTGCAGGATCAGCTGATGAACGAGGAGCTGACGGGAAAGGCTCTTTGTCT 1772

567 uSerAsnSerThrLeuAspAlaSerLeuAlaSerIleAsnLeuGlnArgGlyA 584
    |||||||||||||||||||||
1773 GTCCAATTCCAGCACCTGGATTCGGCGTCCATCAACCTGCAGAGGGGCC 1822

584 rgAspHisGlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeu 600
    ||||||||||||||||
1823 GGGACCACGGCCTGCCAGGTTACAATGAGTGGAGGGAGTTCTGCGGCCTG 1872

601 ProArgLeuGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSe 617
    ||||||||||||||||
1873 CCTCGCCTGGAGACCCCGCTGACCTGAGCACAGCCATCGCCAGCAGGAGA 1922

617 rValAlaAspLysIleLeuLeuAspLysTyrLysHisProAspAsnIleAspV 634
    ||||||||||||||||
1923 CGTGGCCGACAAAGATCCTGGACTTGTACAAGCATCCTGACAACATCGATG 1972

634 alTrpLeuGlyLeuAlaGluAsnPheLeuProArgAlaArgThrGly 650
    ||||||||||||||||
1973 TCTGGCTGGGAGGCTTAGCTGAAAACTTCCTCCCCAGGGCTCGGACAGGG 2022

650 rgLeuPheAlaCysLeuLeuGlyLeuGlnMetLysAlaLeuArgAspGl 667
    ||||||||||||||||
2023 CCCTGTTTGCCTGTCTCATTGGGAAGCAGATGAAGGCTCTGCGGGATGG 2072

667 yAspTrpPheTrpGluAsnSerHisValPheThrAspAlaGlnArgA 684
    ||||||||||||||||
2073 TGACTGGTTTTGGGAGAACAGCCACGTCTTCACGGATGCACAGAGGC 2122

684 rgGluLeuGluLysHisSerSerArgValIleCysAspAsnThrGly 700
    |||||||||||||||||||||||
2123 GTGAGCTGGAAAAGCACTCCTCTCGCGTCATCTGTGACAACACTGGC 2172

701 LeuThrArgValProMetAspAlaPheGlnInValGlyLysPheProGluAs 717
    ||||||||||||||||
2173 CTCACCAGGGTGCCCATGGATGCCTTCCAAGTCGGCAAATTCCCTGAAGA 2222

717 pPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgG 734
    ||||||||||||||||
2223 CTTTGAGTCTTGTGACAGCATCCCTGGCATGAACCTGGAGGCCTGGAGGG 2272

734 luThrPheProGlnAspAspAspLysCysGlyPheProGluSerValGluAsn 750
    ||||||||||||||||
2273 AAACCTTTCCTCAAGACGACAAGTGTGGCTTCCCAGAGAGCGTTGAGAAT 2322

751 GlyAspPheValHisCysGluGluTyrGluValGlyArgArgValLeuValTyrSe 767
    ||||||||||||||||
2323 GGGGACTTTGTGCACTGTGAGGAGTACGAGGTCGGCCGTCGTGTGTTATTC 2372

767 rCysArgHisGlyTyrArgGluLeuInGlyTyrArgGluGlnInLeuThrCysThrG 784
    ||||||||||||||||
2373 CTGCCGGCACGGGTATGAGCTCCAAGGCCGGGAGCAGCTCACTTGCACCC 2422

784 lnGluGlyTrpAspProPheGlnInProProLeuCysLysSerAspValAsnGluCys 800
    ||||||||||||||||
2423 AGGAAGGATGGGATTTCCAGCCTCCCCTCTGCAAAGATGTAAGAGTGT 2472
```

```
801  AlaAspGlyAlaHisProProCysHisAlaSerAlaArgCysArgAsnTh  817
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2473 GCAGACGGTGCCCACCCCCCGCCACGGCCTCTGCGAGGTGCAGAAACAC  2522

817  rLysGlyPheGlnCysLeuCysSerGlyAlaAspProTyrArgLeuGlyAspA  834
     |||||||||||||||||||||||||||||||||||||||||||||||||||||
2523 CAAAGGCGGCTTCCAGTGTCTCTGCGGCGACCCCTACGAGTTAGGAGACG  2572

834  spGlyArgThrCysValAspSerGlyArgLeuProArgValThr  848
     |||||||||||||||||||||||||||||||||||||||:::||
2573 ATGGGAGGAACCTGCGTAGACTCCGGGAGCTCCCTCGGCGACT  2616

seq_name: gb_pat:A92132

seq_documentation_block:
LOCUS       A92132          2847 bp    DNA    circular    PAT        22-JAN-2000
DEFINITION  Sequence 1 from Patent WO9820354.
ACCESSION   A92132
VERSION     A92132.1    GI:6740928
KEYWORDS    .
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2847)
AUTHORS     Furmaniak,J. and Grennan,J.F.
TITLE       MONITORING REACTIVITY OF THYROID PEROXIDASE AUTOANTIBODIES
JOURNAL     Patent: WO 9820354-A 1 14-MAY-1998;
            FURMANIAK JADWIGA (GB); GRENNAN JONES FIONA (GB)
FEATURES             Location/Qualifiers
     source          1..2847
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT    610 a    884 c    833 g    520 t
ORIGIN

alignment_scores:
          Quality: 4525.00            Length:      848
          Ratio:   5.355             Gaps:        0
Percent Similarity: 99.646    Percent Identity: 99.057

alignment_block:
US-08-482-402A-3_COPY_1_848 x A92132
Align seg 1/1  to: A92132   from: 1   to: 2847
```

```
1    MetArgAlaLeuAlaValLeuSerValThrLeuValMetAlaCysThrGl  17
     ||||||||||||||||||||||||||||||||||||||||||||||||||
46   ATGAGAGCGCTCGCTGTCGTGTCTGTCACGCTGGTTATGGCCTGCACAGA  95

17   uAlaPheProPheIleSerArgArgGlyLysGluLeuLeuTrpGlyLysP  34
     ||||||||||||||||||||||||||||||||||||||||||||||||||
96   AGCCTTCCCCTTCATCTCGAGAGGGGAAAGAACTCCTTTGGGGAAAGC  145

34   roGluGluSerArgValSerArgValLeuGluGluSerValLeuSerArgLeuVal  50
     ||||||||||||||||||||||||||||||||||||||||||||||||||
146  CTGAGGAGTCTCGTGTCTCTCAGCCTCTTGGAGGAAAGCAAGCGGCTGGTG  195

51   AspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArgGlyI1  67
     ||||||||||||||||||||||||||||||||||||||||||||||||||
196  GACACCGCCATGTACGCCACGATGCAGAGAAACCTCAAGAAAAGAGGAAT  245

67   eLeuSerGlyAlaGlnLeuLeuSerPheSerLysLysLeuProGluProThrS  84
     ||||||  ||||||||||||||||||||||||||||||||||||||||||
246  CCTTTCCCAGCTCAGCTTCTGTCTTTTTCCAAACTTCCTGAGCCAACAA  295

84   erGlyValIleLeuArgAlaAlaAlaGluI1eMetGluThrSerIleGlnAla  100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
296  GCGGAGTGATTGCCCGAGCGAGAGATAATGGAAACATCAATACAAGCG  345

101  MetLysArgLysValAsnLeuLysThrGlnGlnSerGlnHisProThrAs  117
     ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
346  ATGAAAAGAAAAGTCAACCTGAAAACTCAACATCACAGCATCCAACGGA  395

117  pAlaLeuSerGluAspLeuLeuSerIleIleAlaAsnMetSerGlyCysL  134
     ||||||||||||||||||||||||||||||||||||||||||||||||||
396  TGCTTTATCAGAAGATCTGCTGAGCATCATTGCAAACATGTCTGGATGTC  445

134  euProTyrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLys  150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
446  TCCCTTACGTGCTGCCCCCAAAATGCCAAACACTTGCCTGGCGAACAAA  495

151  TyrArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpG1  167
     ||||||||||||||||||||||||||||||||||||||||||||||||||
496  TACAGGCCCATCACAGGAGCTGCAACAACAGAGACCACCCCAGATGGGG  545

167  yAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspG  184
     ||||||||||||||||||||||||||||||||||||||||||||||||||
546  CGCCTCCAACACCGGCCCTGGCACGATGGCTCCCTCCAGTCTATGAGGACG  595

184  lyPheSerGlnProArgGlyTrpAsnProGlyPheLeuTyrAsnGlyPhe  200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
596  GCTTCAGTCAGCCCCGAGGCTGGAACCCCGGCTTCTTGTACAACGGGTTC  645

201  ProLeuProProValArgGluValThrArgHisValIleGlnValSerAr  217
     ||||||||||||||||||||||||||||||||||||||||||||||||||
646  CCACTGCCCCCGGTCCGGGAGGTGACAAGACATGTCATTCAAGTTCAA  695

217  nGluValThrArgAspAspAspArgTyrSerAspLeuLeuMetAlaTrpG  234
     ||||||||||||||||||||||||||||||||||||||||||||||||||
696  TGAGGTTGTCACAGATGATGACCGCTATTCTGACCTCCTGATGGCATGGG  745

234  lyGlnTyrIleAspHisAspIleAlaPheThrProGlnSerThrSerLys  250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
746  GACAATACATGACCACGACATCGCGTTCACACCACAGAGCACCAGCAAA  795

251  AlaAlaPheGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAs  267
     |||||||||||||||||||||||||:::||||||||||||||||||||||
796  GCTGCCTTCGGGGGAGGGGGCTGACTGCCAGATGACTTGTGAGAACCAAAA  845

267  nProCysPheProIleGlnLeuProGluGlnAlaAlaArgProAlaAlaGlyT  284
     ||||||||||||||||||||||||||||||||||||||||||||||||||
846  CCCATGTTTTCCCATACAACTCCCGAGGAGCCCCGGCCCGGCCGGGGCA  895

284  hrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyThrGlyAsp  300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
896  CGGCCTGTCTGCCCTTCTACCGCTCTTCCGCCGCCTGCGGCACCGGGGAC  945

301  GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe  317
     ||||||||||||||||||||||||||||||||||||||||||||||||||
946  CAAGGCGCGCTCTTTGGGAACCTGTCCACGGCCAACCCGCGGCAGCAGAT  995

317  tAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGlySerSerP  334
     ||||||||||||||||||||||||||||||||||||||||||||||||||
996  GAACGGGTTGACCTCGTTCCTGGACGCGTCCACCGTGATGGCAGCTCCC  1045

334  roAlaLeuGluArgGlnLeuArgAsnTrpThrSerAlaGluGlyLeuLeu  350
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1046 CGGCCCTAGAGAGGCAGCTGCGGAACTGGACCAGTGCCGAAGGGCTGCTC  1095

351  ArgValHisGlyArgLeuArgAspSerGlyArgAlaTyrLeuProPheVa  367
     |||||:::||||||||||||||||||||||||||||||||||||||||||
1096 CGGGTCCACGCGCGCCTCCGGGACTCCGGGCGCGCCTACCTGCCCTTCGT  1145

367  lProProArgAlaProAlaAlaCysAlaProGluProGlyProGlyG  384
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1146 GCCGCCACGCGCGGCCCTGCGGGCCTGCGCCCGAGCCCGGCATCCCGGAG  1195

384  luThrArgGlyProCysPheLeuAlaGlyAspValArgAlaSerGluVal  400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1196 AGACCCCGCGGGCCCTGCTTCCTGGCCGGAGACGGCGCCGCCAGCGAGGTC  1245

401  ProSerLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLe  417
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1246 CCCTCCCTGACGGCACTGCACACGCTGTGGCTGCGCGAGCACAACCGCCT  1295
```

```
717 pPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgG 734
        ||||||||||||||||||  ||||||||||||||||||||||||||||
2196 CTTTGAGTCTTGTGACAGCATCCCTGGCATGAACCTGGAGGCCTGGAGGG 2245

734 luThrPheProGlnAspAspLysCysGlyPheProGluSerValGluAsn 750
        |||||||||||||||||||||||||||||||||||||||:::||||||
2246 AAACCTTTCCTCAAGACAGCAAGTGTGGCTTCCCAGAGACATGGAGAAT 2295

751 GlyAspPheValHisCysGluGluSerGlyArgArgValLeuValTyrSe 767
        ||||||||||||||||||||||||||||||||||||||||||||||||||
2296 GGGGACTTTGTGCACTGTGAGGAGTCTGGGAGGCGCGTGCTGGTGTATTC 2345

767 rCysArgHisGlyTyrGluLeuGlnLysGlyArgGluGlnLeuThrCysThrG 784
        |||||||||||||||| |||||||||||||||||||||||||||||||||
2346 CTGCCGACACGGGTATGAGCTCCAAGGCCGGGAGCCAGAGCTCACTTGCACCC 2395

784 lnGluGlyTrpAspPheGlnProProLeuCysLysSerValAsnGluCys 800
        ||||||||||||||||||||||||||||||||||||||||||||||||
2396 AGGAAGGATGGGATTTCCAGCCTCCCCCTGCAAAGATGTAACCAGTGT 2445

801 AlaAspGlyAlaHisProProCysHisAlaSerAlaArgCysArgAsnTh 817
        |||||||||||||||||||||||||||||||||||||||||||||||||
2446 GCAGACGGTGCCCACCCCCCCTGCCACCGCCTCTGCGAGGTGCAGAAACAC 2495

817 rLysGlyGlyPheGlnCysLeuCysAlaAspProTyrGluLeuGlyAspA 834
        ||||||||||||||||||||||||||||||||||||||||||||||||||
2496 CAAAGGCGGCTTCCAGTGTCTCTGCGCCGACCCCTACGAGTTAGGAGACG 2545

834 spGlyArgThrCysValAspSerGlyArgLeuProArgValThr 848
        ||||||||||||||||||||::::|||||||||||
2546 ATGGGAGAACCTGCGTAGACTCCGGGAGGCTCCCTCGGGGGACT 2589
```

```
seq_name: gb_pr:HSTPO

seq_documentation_block:
LOCUS          HSTPO          3027 bp    mRNA                         PRI      23-MAR-1995
DEFINITION     Human mRNA for thyroperoxidase.
ACCESSION      Y00406
VERSION        Y00406.1 · GI:37250
KEYWORDS       membrane glycoprotein; thyroperoxidase.
SOURCE         human.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1  (bases 1 to 3027)
AUTHORS        Libert,F., Ruel,J., Ludgate,M., Swillens,S., Alexander,N.,
               Vassart,G. and Dinsart,C.
TITLE          Complete nucleotide sequence of the human
               thyroperoxidase·microsomal antigen cDNA
JOURNAL        Nucleic Acids Res. 15 (16), 6735 (1987)
MEDLINE        87316933
REFERENCE      2  (bases 1 to 3027)
AUTHORS        Libert,F.
TITLE          Direct Submission
JOURNAL        Submitted (28-OCT-1987) to the EMBL/GenBank/DDBJ databases
FEATURES             Location/Qualifiers
     source          1..3027
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="thyroid"
                     /clone_lib="lambda gt11"
                     /map="chromosome 2, short arm"
     CDS             41..2842
                     /note="precursor polypeptide"
                     /codon_start=1
                     /protein_id="CAA68467.1"
                     /db_xref="GI:37251"
                     /db_xref="SWISS-PROT:P07202"
                     /translation="MRALAVLSVTLVMACTEAFFPFISRGKELLWGKPEESRVSSVLE
                     ESKRLVDTAMYATMQRNLKKRGILSPAQLLSFSKLPEPTSGVIARAAEIMETSIQAMK
                     RVNLKTQQSOHPTDALSEDLLSIIANMSGCLPYMLPPKCPNTCLANKYRPITGACNN
                     RDHPRWGASNTALARWLPPVYEDGFSQPRGWNPGFLYNGFPLPPVREVTRHVIQVSNE
                     VVTDDDRYSDLLMAWGQYIDHIAFTPQSTSKAAFGGGADCQMTCENQNPCFPIQLPE
```

```
417 uAlaAlaAlaLeuLysAlaLeuAsnAlaHisTrpSerAlaAspAlaValT 434
        |||||||||||||||||||||||||||||||||||||||||||||||||
1296 GGCCCGGCGGCTCAAGGCCCTCAATGCCCACTGGAGCGCCGGACGCCGTGT 1345

434 yrGlnGluAlaArgLysValValGlyAlaLeuHisGlnIleIleThrLeu 450
        |||||||||||||||||||||||||||||||||||||||||||||||
1346 ACCAGGAGGCGCGCAAGGTCGTGGGCGCTCTGCACCAGATCATCACCCTG 1395

451 ArgAspTyrIleLeuProArgIleLeuGlyProGlnAlaPheGlnGlnTyrVa 467
        |||||||||||||||||||||||||||||||||||||||||||||
1396 AGGGATTACATCCCCAGGATCCTGGGACCCCAGGCCTTCCAGCAGTACGT 1445

467 lGlyProTyrGluGlyTyrAspSerThrAlaAsnProThrValSerAsnV 484
        |||||||||||||||||||||||||||||||||||||||||||||||||
1446 GGGTCCCTATGAAGGCTATGACTCCACCGCCAACCCCACTGTGTCCAACG 1495

484 alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu 500
        |||||||||||||||||||||||||||||||||||||||||||||||||
1496 TGTTCTCCACAGCCGCCTTCCGCTTCGGCCACGCCACCATCCACCCGCTG 1545

501 ValArgArgLeuAspAlaSerPheGlnHisGluHisProAspLeuProGlyLe 517
        |||||||||||||||||||||||||||||||||||||||||||||||||
1546 GTGAGGAGGCTGGACGCGTCCAGCTTCCAGGAGCACCCGACCTGCCCGGGCT 1595

517 uTrpLeuHisGlnAlaPhePheSerProTrpThrLeuLeuArgGlyGlyG 534
        |||||||||||||||||||||||||||||||||||||||||||||||||
1596 GTGGCTGCACCAGGCTTTCTTCAGCCCATGGACATTACTCCGGTGGAGGTG 1645

534 lyLeuAspProLeuIleArgGlyLeuLeuAlaArgProAlaLysLeuGln 550
        |||||||||||||||||||||||||||||||||||||||||||||||||
1646 GTTTGACCCCACTAATACGAGGCCTTCTTGCAAGACCCAGCCAAACTGCAG 1695

551 ValGlnAspGlnLeuMetAsnGluGluLeuThrGluArgLeuPheValIle 567
        |||||||||||||||||||||||||||||||||||||||||||||||
1696 GTGCAGGATCAGCTGATGAACGAGGAGCTGACGGAAAGGCTCTTTGTGCT 1745

567 uSerAsnSerThrLeuAspLeuAlaSerIleAsnLeuGlnArgGlyAla 584
        ||||||||||||||||||||::::||||||||||||||||||||||||||
1746 GTCCAATTCCAGCACCTTGAATCTGGCGTCCATCAACCTGCAGAGGGGCC 1795

584 rgAspHisGlyLeuProGlyTyrArgGluTrpArgGluPheCysGlyLeu 600
        |||||||||||||||||||||||||||||||||||||||||||||||||
1796 GGGACCACGGGCTGCCAGTTACAATGAGTGGAGGGAGTTCTGCGGCCTG 1845

601 ProArgLeuGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSe 617
        |||||||||||||||||||||||||||||||||||||||||||||||||
1846 CCTCGCCTGGAGACCCCCGCCGACCTGAGCACAGCCATCGCCAGCAGGAG 1895

617 rValAlaAspLysIleIleLeuAspAspLeuTyrLysHisProAspAsnIleAspV 634
        |||||||||||||||||||||||||||||||||||||||||||||||
1896 CGTGGCCGACAAGATCATCCTGGACGATCTGTACAAGCATCCGACAATCGATG 1945

634 alTrpLeuGlyLeuLeuAlaGluAsnPheLeuProGluArgLeuArgThrGly 650
        |||||||||||||||||||||||||||||||||||||||||||||
1946 TCTGGCTGGGAGGCTTAGCTGAAAACTTCCTCCCCAGAGGCTCGGACCAGGG 1995

651 ProLeuPheAlaCysLeuIleGlyLysGlnMetLysAlaLeuArgAspGl 667
        |||||||||||||||||||||||||||||||||||||||||||||||
1996 CCCCTGTTTGCCTGTCTCATTGGGAAGCAGATGAAGGCTCTGCGGGATGG 2045

667 yAspTrpPheTrpAspMetAspAsnSerHisValPheThrAspAlaGlnArgA 684
        ||||||||||||||||||||||||||||||||||||||||||||||
2046 TGACTGGTTTTGGTGGGAGAACAGCCACGTCTTCACGGATGCACAGAGGC 2095

684 rgGluLeuLysHisSerLeuSerArgValIleCysAspAsnThrGly 700
        ||||||||||||||||||||||||||||||||||||||||||||||
2096 GTGAGCTGGAGAAGCACTCCCTGTCTCGGGTCATCTGTGACAACACTGGC 2145

701 LeuThrArgValProMetAspAlaPheGlnValGlyLysPheProGluAs 717
        ||||||||||||||||||||||||||||||||||||||||||||||||||
2146 CTCACCAGGGTGCCCATGGACGCCTTCCAAGTCGGCAAATTCCCTGAAGA 2195
```

Mon Nov 26 15:43:03 2001   us-08-482-402a-3_copy_1_848.rge

```
EARPAAGTACLPFYRSSAACGTGDQGALFGNLSTANPRQQMNGLTSFLDASTVYGSSP
ALERQLRNWTSAEGLLRVHARLRDSGRAYLPFVPPRAPAACAPEPGIPGETRGPCFLA
GDGRASEVPSLTALHTLWLRERHNRLAAALKALNAHWSADAVYQEARKVVGALHQIITL
RDYIPRILGPEAFQQYVGPYEGYDSTANPTVSNVFSTAAFRFGHATIHPLVRRLDASF
QEHPDLPGLWLHQAFFSPWTLRGGGLDPLIRGLLARPAKLQVDQLMNEELTERLFV
LSNSSTLNLASINLQRGRDHGLPGYNEWREFCGLPRLETPADLSTAIASRSVADKILD
LYKHPDNIDVWLGGLAENFLPRARTGPLFACLLICKQMKALRDGDWFWENSHVFTDAQ
RRELEKHSLSRVICDNTGLTRVPMDAFQVGKFPEDFESCDSIPGMNLEAWRETFPQDD
KGGFPESMENGDFVHCEESGRRVIVYSCRHGYELQGREQLTCTQEGWDFQPPLCKDVN
ECADGANHPCHASARCRNTKGGFQCLCADPYELGDDGRTCVDSGRLPRATWISMSLAA
LLIGGFAGLTSTVICKWTRGTKSTLPISETGGGTPELRCGKHQAVGTSPQRAAAQDS
EQESAGMEGRDTHRLPRAL"
```

sig_peptide   41. .82
                 /product="put. signal peptide (AA -14 to -1)"
mat_peptide   83. .2839
                 /product="put. mature thyroperoxidase (AA 1-919)"
misc_feature   425. .433
                 /note="pot. N-glycosylation site"
misc_feature   959. .967
                 /note="pot. N-glycosylation site"
misc_feature   1064. .1072
                 /note="pot. N-glycosylation site"
misc_feature   1472. .1480
                 /note="pot. N-glycosylation site"
misc_feature   1745. .1753
                 /note="pot. N-glycosylation site"
misc_feature   2579. .2652
                 /note="pot. transmembrane region"
BASE COUNT   664 a   928 c   871 g   564 t
ORIGIN

alignment_scores:
      Quality: 4525.00   Length: 848
      Ratio: 5.355   Gaps: 0
Percent Similarity: 99.646   Percent Identity: 99.057

alignment_block:
US-08-482-402A-3_COPY_1_848 x HSTPO

Align seg 1/1 to: HSTPO from: 1 to: 3027

```
  1 MetArgAlaLeuAlaValLeuSerValThrLeuValMetAlaCysThrGl   17
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 41 ATGAGAGCCCTCGCTGTGCTGTCAGTGACCCTGGTTATGGCCTGCACAGA   90

 17 uAlaPhePheProPheIleSerArgGlyLysLeuLeuTrpGlyLysP   34
    ||||||||||||||||||||||||||||||||||||||||||||||
 91 AGCCCTTCTTCCCCTTCATCTCGAGAGGGAAAGAACTCCTTTGGGGAAAGC  140

 34 roGluGluSerArgValSerSerValLeuGluLeuSerLysArgLeuVal   50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
141 CTGAGGAGTCTCGTGTCTCTAGCGTCTTGGAGGAAAGCAAGCGCCTGGTG  190

 51 AspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArgGlyIl   67
    ||||||||||||||||||||||||||||||||||||||||||||||||||
191 GACACCGCCATGTACGCCACGATGCAGAGAAACTTCAAGAAAAGAGGAAT  240

 67 eLeuSerGlyAlaGlnLeuLeuSerPheSerLysLeuProGluProThrS   84
    ||||||||||||||||||||||||||||||||||||||||||||||||||
241 CCTTTCTCCAGCTCAGCTTCTTTTTCCAAACTTCCTGAGCCCACAA  290

 84 erGlyValIleAlaArgAlaAlaAlaGluIleMetGluThrrserIleGlnAla  100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
291 GCGGAGTGATTGCCCGAGCAGCAGCAGATAATGGAAACATCAATACAAGCG  340

101 MetLysArgLysValAsnLeuLysThrGlnGlnSerGlnHisProThrAs  117
    ||||||||||||||||||||||||||||||||||||||||||||||||||
341 ATGAAAAGAAAAGTCAACCTGAAAACTCAACAATCACAGCATCCAACGA  390

117 pAlaLeuSerGluAspLeuLeuSerIleIleAlaAsnMetSerGlyCysL  134
    |||||||||||||||||||||||||||||||||||||||||||||||||
391 TGCTTTATCAGAAGATCTGCTGAGCATCATTGCAAACATGTCTGGATGTC  440
```

                                                                    euProTyrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLys 150
                                                                134 |||||||||||||||||||||||||||||||||||||||||||||||||||
                                                                441 TCCCTTACATGCTGCCCCCAAAATGCCCAAACACTTGCCTGGCGAACAAA 490

                                                                    TyrArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpGl 167
                                                                151 |||||||||||||||||||||||||||||||||||||||||||||||||||
                                                                491 TACAGGCCCATCACAGGAGCTTGCAACACAGAGACCACCCCAGATGGGG 540

                                                                    yAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspG 184
                                                                167 |||||||||||||||||||||||||||||||||||||||||||||||||||
                                                                541 CGGCTCCCAACACGGCCCTGGCACGATGGCTCCCTCCAGTCTATGAGGACG 590

                                                                    lyPheSerGlnProArgGlyTrpAsnProGlyPheLeuTyrAsnGlyPhe 200
                                                                184 |||||||||||||||||||||||||||||||||||||||||||||||||||
                                                                591 GCTTCAGTCAGCCCCGAGGCTGGAACCCCGGCTTCTTGTACAACGGGTTC 640

                                                                    ProLeuProProValArgGluValThrArgHisValIleGlnValSerAs 217
                                                                201 |||||||||||||||||||||||||||||||||||||||||||||||||||
                                                                641 CCACTGCCCCCGGTCCGGGAGGTGACAAGACATGTCATTCAAGTTTCAAA 690

                                                                    nGluValThrAspAspAspArgTyrSerAspLeuLeuMetAlaTrpG 234
                                                                217 ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                                691 TGAGGTTGTCACAGATGATGACCGCTATTCTGACCTCCTGATGGCATGGG 740

                                                                    lyGlnTyrIleAspHisAspIleAlaPheThrProGlnSerThrSerLys 250
                                                                234 |||||||||||||||||||||||||||||||||||||||||||||||||||
                                                                741 GACAATACATCGACCACGACATCGCGTTCACACCACAGAGCACCAGCAAA 790

                                                                    AlaAlaPheGlyGlyGlySerAspCysSerGlnMetThrCysGluAsnGlnAs 267
                                                                251 ||||||||||||||||:::||||||||||||||||||||||||||||||||
                                                                791 GCTGCCTTCGGGGAGGGCTGACTGCCAGATGACTTGTGAGAACCAAAA 840

                                                                    nProCysPheGlnProIleGlnLeuProGluGluAlaArgProAlaAlaGlyT 284
                                                                267 ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                                841 CCCATGTTTTCCCATACAACTCCCGGAGGAGCCCGGCCGGCCGGGCA 890

                                                                    hrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyThrGlyAsp 300
                                                                284 ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                                891 CCGGCCGTGCTGCCCCTTCTACCGCTCTTCGGCCGCCACCGGGGAC 940

                                                                    GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe 317
                                                                301 ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                                941 CAAGGCGGCTCTTTGGGAACCTGTCCACGGCCAACCCGCGGCAGCAGAT 990

                                                                    tAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGlySerSerP 334
                                                                317 ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                                991 GAACGGGTTGACCTCGTTCCTGGACGCGTCCACCGTGTATGGCAGCTCCC 1040

                                                                    roAlaLeuGluArgGlnLeuArgAsnTrpThrSerAlaGluGlyLeuLeu 350
                                                                334 ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                                1041 CGGCCCTAGAGAGGCAGCTGCGGAACTGGACCAGTGCCGAAGGGCTGCTC 1090

                                                                    ArgValHisGlyArgLeuArgAspSerGlyArgAlaTyrLeuProPheVa 367
                                                                351 ||||||||||||:::|||||||||||||||||||||||||||||||||||
                                                                1091 CGGGTCCACGGCGCCCTCCGGGACTCCGGGCGCGCCTACCTGCCCTTCGT 1140

                                                                    lProProArgAlaProAlaAlaCysAlaProGluProGlyAsnProGlyG 384
                                                                367 ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                                1141 GCCGCCACGGGCGCCGGCCGCCTGCGCCCCGGAGCCCGGGCATCCCCGGAG 1190

                                                                    luThrArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerGluVal 400
                                                                384 |||||||||||||||||||||||||||||||||||||||||||||||||||
                                                                1191 AGACCCGCGGGCCCTGCTTCCTGGCCGGCGACGGCCGCCAGCGAGGTC 1240

                                                                    ProSerLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLe 417
                                                                401 |||||||||||||||||||||||||||||||||||||||||||||||||||
                                                                1241 CCCTCCCTGACGGCACTGCACACGCTGTGGCTGCGCGAGCACAACCGCCT 1290

                                                                    uAlaAlaAlaLeuLysAlaLeuAsnAlaHisTrpSerAlaAspAlaValT 434
                                                                417 ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                                1291 GGCCGCGGCGCTCAAGGCCCTCAATGCGCACTGGAGCGCGGACGCCGTGT 1340

```
434  yrGlnGluAlaArgLysValValGlyAlaLeuHisGlnIleIleThrLeu  450
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1341 ACCAGGAGGCGCAAGGTCGTGGGCGCTCTGCACCAGATCATCACCCTG     1390

451  ArgAspTyrIleProArgIleLeuGlyProGluAlaPheGlnGlnTyrVa   467
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1391 AGGGATTACATCCCCAGGATCCTGGGACCCGAGGCCTTCCAGCAGTACGT   1440

467  lGlyProTyrGluGlyTyrArgAspSerThrAlaAsnProThrValSerAsnV  484
     |||||||||||||||||||||||||||||||||||||||||||||||||||||
1441 GGGTCCCTATGAAGGCTATCGACTCCACCGCCAACCCCACTGTGTCCAACG  1490

484  alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu    500
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1491 TGTTCTCCACAGCCGCCTTCCGCTTCGGCCATGCCACGATCCACCCGCTG   1540

501  ValArgLeuAspAspAlaSerPheGlnGluHisProAspLeuProGlyLe    517
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1541 GTGAGGAGCGGCTGGAGCACGCCAGCTTCCAGGAGCACCCCGACCTGCCCGGCT  1590

517  uTrpLeuHisGlnAlaPheSerProProThrPheLeuArgGlyGlyG      534
     |||||||||||||||||||||||||||||||||||||||||||||||||
1591 GTGGCTGCACCAGGCTTTCTTCCAGCCATGGACATTACTCCGTGAGGTG   1640

534  lyLeuAspProArgLeuIleArgGlyLeuLeuAlaArgProAlaLysLeuGln  550
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1641 GTTTTGGACCCCACTAATACGAGGCCTTCTTGCAAGACCAGCCAAACTGCAG  1690

551  ValGlnAspGlnLeuMetAsnGluLeuThrGluLeuArgLeuPheValLe     567
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1691 GTGCAGGATCAGCTGATGAACGAGGAGCTGACGGAGCTCTTTGTGCT      1740

567  uSerAsnSerThrLeuAspLeuAlaSerIleAsnLeuGlnArgGlyA      584
     |||||||||||:::|||||||||||||||||||||||||||||||||||
1741 GTCCAATTCCAGCACCTTGAATCTGGCGTCCATCAACCTGCAGAGGGGCC   1790

584  rgAspHisGlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeu    600
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1791 GGGACCACGGCCTGCCAGGTTACAATGAGTGGAGGGAGTTCTGCGGCCTG  1840

601  ProArgLeuGluThrProAlaAspPheLeuSerThrAlaIleAlaSerArgSe  617
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1841 CCTCGCCTGGAGACCCCGCTGACTTCACGCACAGCCATCGCCAGCAGGAG  1890

617  rValAlaAspLysIleIleLeuLeuTyrLysHisProAspAsnIleAsppV    634
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1891 CGTGGCCGACAAGATCATCCTGCTGTACAAGCATCCTGACAACATCGATG  1940

634  alTrpLeuGlyGlyLeuAlaGluAsnPheLeuProArgAlaArgThrGly    650
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1941 TCTGGCTGGGAGGCTTAGCTGAAAACTTCCTCCCCAGGGCTCGGACAGGG   1990

651  ProLeuPheAlaCysLeuIleGlyLysGlnMetLysAlaLeuArgAspGl     667
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1991 CCCCTGTTTGCCTGTCTCATTGGGAAGCAGATGAAGGCTCGCGGGATGG    2040

667  yAspTrpPheTrpTrpGluAsnSerHisValPheThrAspAlaGlnArgA    684
     |||||||||||||||||||||||||||||||||||||||||||||||||||
2041 TGACTGGTTTTGGTGGGAGAACAGCCACGTCTTCACGGATGCACAGAGGC   2090

684  rgGluLeuGluLysHisSerLeuSerArgValIleCysAspAsnThrGly    700
     |||||||||||||||||||||||||||||||||||||||||||||||||||
2091 GTGAGGAGAAGCATTCCCTGTCTCGGGTCATCTGCGACAACACTGGC     2140

701  LeuThrArgValProMetAspAlaPheGlnValGlyLysPheProGluA     717
     |||||||||||||||||||||||||||||||||||||||||||||||||||
2141 CTCACCAGGGTGCCCATGGATGCCTTCCAAGTCGGCAAATTCCCTGAAGA  2190

717  pPheGluSerSerIleThrGlyMetAsnLeuAlaTrpArgG            734
     |||||||||||||||||||||||||||||||||||||||||||
2191 CTTTGAGTCTTGTGACAGCATCCCTGGCATGAACCTGGAGGCCTGGAGGG  2240

734  luThrPheProGlnAspLysCysGlyPheCysGlyPheGluSerValGluAsn  750
```

```
2241 AAACCTTCCTCAAGACGACAAGTGTGGCTTCCAGAGACATGGAGAAT      2290
     ||||||||||||||||||||||::::|||||||

751  GlyAspPheValHisCysGluCysGluSerGlyArgArgValLeuValTyrSe   767
     |||||||||||||||||||||||||||||||||||||||||||||||||||||
2291 GGGGACTTTGTGCACTGTGAGTGTGAGAGTGGCCGCCGTGTGGTGTATTC  2340

767  rCysArgHisGlyTyrGluGlyLeuGlnLeuGlnGlyLeuThrCysThrG   784
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2341 CTGCCGGCACGGGTATGAGCTCCAAGGCCGGGAGCAGCTCACTTGCACCC   2390

784  lnGluGlyTrpAspPheGlnProLeuCysLysAspValAsnGluCys      800
     ||||||||||||||||||||||||||||||||||||||||||||||
2391 AGGAAGGATGGGATTTCCAGCCCCTGTGCAAAGATGTGAACGAGTGT     2440

801  AlaAspGlyAlaHisProProCysHisAlaSerAlaArgCysArgAsnTh    817
     |||||||||||||||||||||||||||||||||||||||||||||||||||
2441 GCAGACGGTGCCCACCCCCCTGCCACGCCTCTGCGAGGTGCAGAAACAC   2490

817  rLysGlyGlyTyrLeuCysSerAlaAspProTyrGluLeuGlyLeuGlyAspA  834
     |||||||||||||||||||||||||||||||||||||||||||||||||||||
2491 CAAAGGCGGCTTCCAGTGTCTCTGCGGCGACCCCTACGAGTTAGGAGACG  2540

834  spGlyArgThrCysValAspSerSerGlyArgLeuProArgArgValThr    848
     ||||||||||||||||||||||::::||||||
2541 ATGGGAGACGGACCTGCGTAGACTCCGGGAGGCTCCCTCGGGCGACT     2584
```

seq_name: gb_pat:E23825

```
seq_documentation_block:
LOCUS       E23825       2546 bp     DNA                    PAT       07-FEB-2001
DEFINITION  Antigen for immunoassaying antihuman thyroid peroxidase antibody
            and recombinant human thyroid peroxidase.
ACCESSION   E23825
VERSION     E23825.1    GI:13024570
KEYWORDS    JP 199094833-A/1.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified
REFERENCE   1  (bases 1 to 2546)
  AUTHORS   Masao,F.
  TITLE     Antigen for immunoassaying antihuman thyroid peroxidase antibody
            and recombinant human thyroid peroxidase
  JOURNAL   Patent: JP 199094833-A 1 09-APR-1999;
            SRL INC
COMMENT     OS   Unidentified
            PN   JP 199094833-A/1
            PD   09-APR-1999
            PF   19-SEP-1997 JP 1997273743
            PR
            PI   MASAO FUKUSHIMA
            PC   G01N33/53,C12N9/08,C12N15/09,G01N33/531,G01N33/564//(C12N9/08,
            PC   G01N33/53,C12N9/08,C12R1:91),
            PC   C12R1:91),
            PC   C12N15/00
            CC   Strandedness: Double;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   source          1..2546
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1..2546
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      542 a    804 c    725 g    475 t
ORIGIN
```

```
alignment_scores:
               Quality:    4495.00          Length:      848
               Ratio:         5.338         Gaps:          0
Percent Similarity:  99.292    Percent Identity:   98.349

alignment_block:
```

us-08-482-402a-3_copy_1_848.rge

```
US-08-482-402A-3_COPY_1_848 x E23825

Align seg 1/1   to: E23825   from: 1   to: 2546

    1 MetArgAlaLeuAlaValAlaLeuSerValThrLeuValMetAlaCysThrGl  17
      |||||||||||||||||||||||||||||||||||||||||||||||
    1 ATGAGAGCGCTGGCTGTGTCTGTCACGCTGGTTATGGCCTGCACAGA        50

   17 uAlaPhePheProIleSerArgGlyLysGluLeuLeuTrpGlyLysP       34
      ||||||||||||||||||||||||||||||||||||||||||||||
   51 AGCCTTCTTCCCCTTCATCTCGAGAGGGAAAGAACTCCTTTGGGGAAAGC    100

   34 roGluGluSerArgValSerSerValLeuGluSerLysArgLeuVal       50
      |||||||||||||||||||||||||||||||||||||||||||||||
  101 CTGAGGAGTCTCGTGTCTTCAGTCCTTGGAGAAAAGCAAGCCGCTGGTG    150

   51 AspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArgGlyIl    67
      ||||||||||||||||||||||||||||||||||||||||||||||||
  151 GACACCGCCATGTACGCCACGATGCAGAGAAACCTCAAGAAAAGAGGAAT   200

   67 eLeuSerGlyAlaGlnLeuLeuSerPheSerLysLeuProGluProThrS    84
      ||||||||||||||||||||||||||||
  201 CCTTTCTCCAGCTCTCGTCTTTTTCCAAACTTCCTGAGCCAACAA       250

   84 erGlyValIleLeuAlaArgAlaAlaGluIleMetGluThrSerIleGlnAla 100
      |||||||||||||||||||||||||||||||||||||||||||||||||
  251 GCGGAGTGATTGCCCGAGCAGCAGAGATAATGGAAACATCAANTACAAGCG  300

  101 MetLysArgLysValAsnLeuLysThrGlnInSerGlnInHisProThrAs 117
      |||||||||||||||||||||||||||||||||||||||||||||||||
  301 ATGAAAAGAAAGTCAACCTGAAAACTCAACAATCACAGCATCCAACCGA   350

  117 pAlaLeuSerGluLeuSerIleIleAlaAsnMetSerGlyCysL       134
      |||||||||||||||||||||||||||||||||||||||||
  351 TGCTTTATCAGAAGATCTGCTGAGCATCATTGCAAACATGTCTGGATGTC  400

  134 euProThrTyrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLys 150
      |||||||||||||||||||||||||||||||||||||||||||||
  401 TCCCTTACATGCTGCCCCAAAATGCCCAACACTTGCCTGGCGAACAAA    450

  151 TyrArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpGl 167
      ||||||||||||||||||||||||||||||||||
  451 TACAGGCCCCATCACAGAGGCTTGCAACAACAGAGACCACCCAGATGGGG  500

  167 yAlaSerAsnThrAlaAlaLeuAlaArgTrpLeuProProValTyrGluAspG 184
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  501 CGCCTCCAAACACGGCCGTGGCTGCACGATGGCTCCCTCCAGTCTATGAGGACG 550

  184 lyPheSerGlnProIleAlaPheAlaPheGlyTrpAsnGlyPhe       200
      |||||||||||||||||||||||||||||||||||||
  551 GCTTCAGTCAGCCCATCGCATTCGCGTTCACACCAGGTTGTACAACGGGTTC 600

  200 lyGlnTyrIleAspHisAspCysGlnHisTyrGluIlePheProGlyIleAs 217
      |||||||||||||:::||||||||||||||||||||||||
  601 CCACTGCCCCCCGGTCCGGGAGGTGACAAGACATGTCATTCAAGTTCAAA  650

  217 nGluValValThrAspAspAspArgTyrSerAspArgTyrSerAspLeuMetAlaTrpG 234
      |||||||||||||||||||||||||||||||||||||||||||||||
  651 TGAGGTTGTCACAGATGACGCCTATTCTGACCTCCTGATGGCATGGG     700

  234 lyGlnTyrIleAspHisAspCysGlnHisTyrSerLys           250
      |||||||||||||||||||||||||||||||||||||||
  701 GACAATACATCGACCACGACATCGCGTTCACACCACCAGCAGCCAAA     750

  251 AlaAlaPheGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAs 267
      ||||||||||||||||||:::||||||||||||||||||||||
  751 GCTGCCCTTCGGGGAGGGGGGCGACTTGCCAGGATGACTTGTGAGAACCAAA 800

  267 nProCysPheProIleGlnLeuProGluGluGlyAlaAlaArgProAlaAlaAlaGlyT 284
      ||||||||||||||||||||||||||||||||||||||||||||||||
  801 CCCATGTTTTCCATACAACTCCCGGAGGAGGCCCCGCCCGCGGCA      850

  284 hrAlaCysLeuProPheTyrArgSerArgSerAlaAlaCysGlyThrGlyTyrAsp    300
      |||||||||||||||||||||||||||||||||||||||||||||||||
```

```
      |||||||||||||||||||||||||||||||||||||||||||||||||
  851 CGGCCCTGTCTGCCCTCTTCTACCGCTGCTTCTGCGGCCACCGGTAC    900

  301 GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe   317
      ||||||||||||||||||||||||||||||||||||||||||||||
  901 CAAGGCGGCCTCTTGGAAACCTGTCCACGCCAACCCGCGCCAGCAGAT    950

  317 tAsnGlyLeuThrSerPheLeuAspSerProAlaSerThrValTyrGlySerSerP 334
      ||||||||||||||||||||||||||||||||||||||||||||||
  951 GAACGGGTTGACCTCGTTCCTGGACGCGTCCACCGTGATGGCAGCTCCC   1000

  334 roAlaLeuGluArgGlnLeuArgAsnTrpThrSerArgAlaGluLysLeuLeu   350
      ||||||||||||||||||||||||||||||||||||||||||||||
 1001 CGGCCCTAGAGAGGCAGCTGCGAAACTGGACCAGTGCCGAAGGGCTGCTC  1050

  351 ArgValHisGlyArgLeuArgAspSerGlyArgAlaTyrLeuProPheVa   367
      |||||||||||::::||||||||||||||||||||||||||||
 1051 CGCGTCCACGCGCGCCTTCCGGGACTCCGGCCGCGCCCTACCTGCCCTTCGT 1100

  367 lProProArgAlaProGluAlaCysAlaProGluProGlyAsnProGlyG   384
      ||||||||||||||||||||||||||||||||||||||||||||||||
 1101 GCGCCCCACGCGCGCCCTTCCGCCCTGTGCGCCCGGAGCCCGGCATCCCCGGAG 1150

  384 luThrArgGlyProCysPheLeuAlaGlyAspGlyArgGlaAlaSerGluVal  400
      |||||||||||||||||||||||||||||||::::|||||||||||
 1151 AGACCCCGCCGGGCCCTGCTTCCTGGCCGGCGGAGACGGCCGCCCACCGAGGTC 1200

  401 ProSerLeuThrArgAlaLeuHisThrLeuTrpLeuArgGluArgGluHisAsnArgLe 417
      |||||||||||||||||||||||||||||||||||||||
 1201 CCCTCCCTGACGCGCACTGCCACGGCTGTGGCCTGCGGAGCACAACCGCCT  1250

  417 uAlaAlaAlaLeuLysAlaLeuAsnAlaHisSerTrpSerArgAlaAspAlaValT  434
      ||||||||||||||||||||||||||||||||||||||||
 1251 GGCCGCGGCGCTCAAGGCCCTCAATGCGCACTCGTGGAGCCGGGCGGACGCCGTGT 1300

  434 yrGlnGluAlaArgLysValValGlyAlaLeuHisGlnIleIleThrLeu   450
      ||||||||||::::||||||||||||||||||||||||::::|||
 1301 ACCAAGATGCGGCGCAAGGTCGTGGGCGCTCTGCACCAGATCATCACCCTT  1350

  451 ArgAspTyrIleProArgLeuGlyTyrIleAlaProGlnGlnInTyrVa   467
      |||||||||||||||||||::::|||||||||||||||||||||
 1351 AGGGATTACATCCCCAGGCTCGGATACATCGCGCCTTCCAGGCAGTAGCGT 1400

  467 lGlyProTyrGlyGlyTyrAspSerProThrAlaAsnProThrValSerAsnV 484
      |||||||||||||||||||||||||||||||||||||||||||
 1401 GGGTCCCATGAGAGGCTATGACTCACCCACGCCCAACCCACTGTGTCCAACG 1450

  484 alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu   500
      ||||||||||||||||||||||||||||||||||||||||||||
 1451 TGTTCTCCACAGCCGCCTTCCGCTTCGGCCATGCCACGATCCACCCCGCTG  1500

  500 ValArgArgLeuAspAlaSerPheGlnGluGlnLeuThrGluValArgLeuValLe  517
      |||::::|||||||||||||||||||||||||||||||||||||
 1501 GTGAGGAGGCTGGACGCCAGCTTCCAGGAGCAGCTGACGGAAGGCTCTTTGTGCT 1550

  517 uTrpLeuHisGlnAlaPheGlnHisSerProThrLeuLeuArgGlyGlyG   534
      ||||||||||||||||||||||||||||||||||||||||
 1551 GTGGCTGCACCAGGCTTCTTCAGCCCATGGACAATTACTCCGTGGAGGTG  1600

  534 lyLeuAspProLeuIleArgGlyLeuLeuAlaArgProAlaLysLeuGln   550
      |||||||||||||||||||||||||||||||||||||||||
 1601 GTTTGGGACCCACTAATACGAGGCCTTCTTCGAAGGACCAGCCAAACTGCAG 1650

  550 ValGlnAspGlnLeuMetAsnGluLeuThrGluArgLeuPheValLe     567
      |||||||||||||||||||||||||||||||||||||||||
 1651 GTGCAGGATCAGCTGATGAACGAGCTGACGGAGAGGCTCTTTGTGCT    1700

  567 uSerAsnSerSerThrLeuAspSerLeuAlaSerIleAsnLeuGlnArgGlyA  584
      |||||||||||||||||||||||||||||||||||||||||||||
 1701 GTCCAATTCCAGCACCCTTGGATCTGCCGTCCATCAACCTGCAGAGGGGGCC  1750

  584 rgAspHisGlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeu   600
      ||||||||||||||||||||||||||||||||||||||||||||
```

```
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2845)
AUTHORS     Barnett,P.S.
TITLE       Direct Submission
JOURNAL     Submitted (20-DEC-1989) Barnett P.S., King's College School of
            Medicine & Dentistry, Denmark Hill, London SE5 8RX, United Kingdom
REFERENCE   2 (bases 1 to 2845)
AUTHORS     Barnett,P.S., Banga,J.P., Watkins,J., Huang,G.C., Gluckman,D.R.,
            Page,M.J. and McGregor,A.M.
TITLE       Nucleotide sequence of the alternatively spliced human thyroid
            peroxidase cDNA, TPO-2
JOURNAL     Nucleic Acids Res. 18 (3), 670 (1990)
MEDLINE     90175008
FEATURES             Location/Qualifiers
     source          1..2845
                     /organism="Homo sapiens"
                     /strain="Graves' patient"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /tissue_type="thyroid"
                     /clone_lib="lambda gt10"
                     /clone="TPO-21"
     CDS             66..2696
                     /note="thyroid peroxidase (AA 1-876)"
                     /codon_start=1
                     /protein_id="CAA35235.1"
                     /db_xref="GI:28898"
                     /db_xref="SWISS-PROT:P07202"
                     /translation="MRALAVLSVTLVMACTEAFFPISRGKELLWGKPEESRVSSVLE
                     ESKRLVDTAMYATMQRNLKKRGILSPAQLLSFSKLPEPTSGVIARAAEIMETSIQAMK
                     RKVNLKTQQSOHPTDALSEDLLSIIANMSGCLPYMLPPKCPNTCLANKYRPITGACNN
                     RDHPRWGASNTALARWLPPVYEDGFSOPRGWNPGFLYNGFPLPPVREVTRHVIQVSNE
                     VVTDDDRYSDLLMAWGQYIDHDIAFTPQSTSKAAFGGGADCOMTCENQNPCFPIQLPE
                     EARPAAGTACLPFYRSSAACGTDQGALFGNLSTANPRQQMNGLTSFLDASTVYGSSP
                     ALERQLRNWTSAEGLLRVHARLRDSGRAYLPFVPPRAPSACAPEPGIPGETRGPCFLA
                     GDGRATEVPSLTALHTLWLREHNRLAAALKALNAHWSADAVYQEARKVVGALHQIITL
                     RDYIPRILGPEAFQQYVGPYEGYDSTANPTVSNVFSTAAFRFGHATIHPLVRRLDASF
                     QEHPDLPGLWLHQAFFSPWTLLRGGYNEWREFCGLPRLETPADLSTAIASRSVADKIL
                     DLYKHPDNIDVWLGGLAENFLPRARTGPLFACLIGKQMKALRDGDWFWENSHVFTDA
                     QRRELEKHSLSRVICDNTGLTRVPMDAFQVGKFPEDFESCDSITGMNLEAWRETFPQD
                     DKCGFPESVENGDFVHCEESGRRVLVYSCRHGYELQGREQLTCTQEGWDFQPPLCKDV
                     NECADGAHPPCHASARCRNTKGFQCILCADPYELGDDGRTCVDSGRLPRVTWISMSLA
                     ALLIGGFAGLTSTVICWRTRTGKSTLPISETGGTPELRCGKHQAVTSPQRAAAQD
                     SEQESAGMEGRDTHRLPRAL"
BASE COUNT      622 a    880 c    819 g    524 t
ORIGIN
```

```
alignment_scores:
     Quality:   4215.50               Length:  848
     Ratio:        5.343              Gaps:      1
Percent Similarity:   93.042      Percent Identity:  92.571

alignment_block:
US-08-482-402A-3_COPY_1_848 x HSASTPO2

Align seg 1/1  to: HSASTPO2    from: 1  to: 2845

   1 MetArgAlaLeuAlaValLeuSerValThrLeuValMetAlaCysThrGl  17
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  66 ATGAGAGCCCTGGCTGTCCTGTCCGTCACGCTGGTTATGGCCTGCACAGA 115

  17 uAlaPhePheProPheIleSerArgGlyLysGluLeuTrpGlyLysP  34
     |||||||||||||||||||||||||||||||||||||||||||||||
 116 AGCTTTCTTCCCCTTCATCTCGAGAGGGAAAGAACTCCTTTGGGGAAAGC 165

  34 roGluGluSerArgValSerSerValLeuGluSerLysArgLeuVal  50
     ||||||||||||||||||||||||||||||||||||||||||||||||
 166 CTGAGGAGTCTCGTGTCTCTAGCCGTCTTGGAGAAAGCAAGCGCCTGGTG 215

  51 AspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArgGlyIl  67
```

```
1751 GGGACCACGGCGCTGCCAGTTACAATGAGTGGAGGAGTTCTGCGGCCTG 1800
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 601 ProArgLeuGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSe 617

1801 CCTCGCCTGGAGACCCCCGACCTGAGCACAGCCATCGCCAGCAGGAG 1850
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 617 rValAlaAspLysIleLeuAspLeuTyrLysHisProAspAsnIleAspV 634

1851 CGTGGCCGACAAGATCCTGACTTGTACAAGCATCCTGACAACATCGATT 1900
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 634 alTrpLeuGlyGlyLeuAlaGluAsnPheLeuProArgAlaArgThrGly 650

1901 TCTGGCTGGGAGGCTTAGCTGAAAACTTCCTCCCCAGGGCTCGGACAGGG 1950
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 651 ProLeuPheAlaCysLeuIleGlyLysGlnMetLysAlaLeuArgAspGl 667

1951 CCCCTGTTTGCCTGTCTCATTGGGAAGCAGAATGAAGGCTCTGCGGACGG 2000
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 667 yAspTrpPheTrpGluAsnSerHisValPheThrAspAlaGlnArgArgA 684

2001 TCACTGGTTTGGTGGGAGAACAGCCACGTCTTCACGGATGCACAGAGGC 2050
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 684 rgGluLeuGluLysHisSerLeuSerArgValIleCysAspAsnThrGly 700

2051 GTGAGCTGGAAGCACTCCCTGTTCTCGGGTCATCTGTGACAACACTGGC 2100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 700 LeuThrArgValProMetAspAlaPheGlnValGlyLysPheProGluAs 717

2101 CTCACCAGGTGCCCATGGACGCCTTCCAAGTCGGCAAATTCCCCGAAGA 2150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 717 pPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgG 734

2151 CTTTGAGTCTTGTGACAGATCATGGCATGAACCTGGAGGCCTGGAGGG 2200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 734 luThrPheProGlnAspAspLysCysGlyPheProGluSerValGluAsn 750

2201 AAACCTTTCCTCAAGACGACAAGTGTGGCTTCCCAGAGAGCGTGGAGAAT 2250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 750 GlyAspPheValHisCysGluGluSerGlyArgArgValLeuValTyrSe 767

2251 GGGGACTTTGTGCACTGTGAGGAGTCTGGGAGGCGCGTGCTGGTGTATTC 2300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 767 rCysArgHisGlyTyrGluLeuGlnGlyArgGluGlnLeuThrCysThrG 784

2301 CTGCCGACACGGGGATGAGCTCCAAGGCCGGGAGCAGCTCACTTGCACCC 2350
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 784 lnGluGlyTrpAspPhePheGlnProProLeuCysLysAspValAsnGluCys 800

2351 AGGAAGGATGGGATTTCCAGCCTCCCCTCTGCAAAGATGTGAACGAGTGT 2400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 801 AlaAspGlyAlaHisProProCysHisAlaSerAlaArgCysArgAsnTh 817

2401 GCAGACGGTGCCACCCCCCCCCCGCCTCTGCAGTGCAGAAACAC 2450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 817 rLysGlyPheGlnCysLeuCysAlaAspProTyrGluLeuGlyAspA 834

2451 CAAAGGCCGGCTCCCAGTGTCTCTGCCCGACCCCTACGAGTTAGGAGACG 2500
     ||||||||||||||||||||||::::||
 834 spGlyArgThrCysValAspSerGlyArgLeuProArgValThr 848

2501 ATGGGAGGAACCTGCGTAGACTCCGGGAGGCTCCCTCGGGCGACT 2544
```

```
seq_name: gb_pr:HSASTPO2

seq_documentation_block:
LOCUS       HSASTPO2     2845 bp    mRNA           PRI       12-SEP-1993
DEFINITION  Human mRNA from alternatively spliced TPO-2 gene for thyroid
            peroxidase (EC.1.11.1.7).
ACCESSION   X17358
VERSION     X17358.1  GI:28897
KEYWORDS    autoantigen; peroxidase; thyroid peroxidase; TPO-2 gene;
            transmembrane protein.
SOURCE      human.
```

```
216 |||||||||||||||||||||||||||||||||||||||||||||||||| 265
    GACACCGCCATGTACGCCACGATGCAGAGAGAAACCTCAAGAAAAGAGGAAT

 67 eLeuSerGlyAlaGlnLeuLeuSerPheSerLysLeuProGlyProThrS  84
    |||||||||||||||||||||||||| |||||||||||||
266 CCTTTCTCCAGCTTCTGTCTTTTTCCAAACTTCCTGAGCCAACAA       315

 84 erGlyValIleAlaAlaGlyAlaAlaGluIleMetGluThrSerIleGlnAla 100
    ||||||||||||||||||||||||||||||::::||||||||
316 GCGGAGTGATTGCCCGAGCAGCAGAGAGATAATGGAAACATCAATACAAGCG 365

101 MetLysArgLysValAsnLeuLysThrGlnGlnSerGlnHisProThrAs  117
    ||||||||||||||||||||||||||||||||||||||
366 ATGAAAAGAAAAGTCAACCTGAAAACTCAACAATCACAGCATCCAACGGA   415

117 pAlaLeuSerGluLeuSerIleIleAlaAsnMetSerGlyCysL        134
    ||||||||||||||||||||||||
416 TGCTTTATCAGAAGATCTGCTGAGCCATCATTGCAAACATGTCTGGATGTC 465

134 euProTyrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLys  150
    ||||||||||||||||||||||||
466 TCCCTTACATGCTGCCCCCAAAATGCCCAAACACTTGCCTGGCGAACAAA  515

151 TyrArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpGl  167
    ||||||||||||||||||||||||
516 TACAGGCCCCATCACAGGAGCTTGCAACACAGAGACCACCCCAGATGGGG  565

167 yAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspG  184
    ||||||||||||||||||||
566 CGGCTCCAACACGGCCCTGGCACGATGGCTCCCAGTCTATGAGGACG     615

184 lyPheSerGlnProArgGlyTrpAsnProGlyPheLeuTyrAsnGlyPhe  200
    |||||||||||||||||||
616 GCTTCAGTCAGCCCCGAGGCTGGAACCCCGGCTTCTTGTACAACGGGTTC  665

201 ProLeuProProValArgGluValThrArgHisValIleGlnValSerAs  217
    |||||||||||||||||||||
666 CCACTGCCCCCGGTTCGGGAGGTGACAAGACATGTCATTCAAGTTTCAAA  715

217 nGluValValThrArgAspAspArgTyrSerAspLeuLeuMetAlaTrpG  234
    ||||||||||||||||||||
716 TGAGGTTGTCACAGATGATGACCGCTATTCTGACCTCCTGATGGCATGGG  765

234 lyGlnTyrIleAspHisAspIleAlaPheThrProGlnSerThrSerLys  250
    ||||||||||||||||||
766 GACAATACATCGACCACGACATCGCGTTCACACCACAGAGCACCAGCAAA  815

251 AlaAlaPheGlyGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAs 267
    |||||||||||||||||||
816 GCTGCCTTCGGCGGGGGCTCTGACTGCCAGATGACTTGTGAGAACCAAAA  865

267 nProCysPheProIleGlnLeuProGluValAlaArgProAlaAlaGlyT  284
    ||||||||||||||||||
866 CCCATGTTTTCCCATACAACTCCCGAGGAGGCCCGGCCCGCCGCGGGCA   915

284 hrAlaCysLeuProTyrArgSerSerAlaAlaCysGlyThrGlyAspAsp  300
    ||||||||||||||||||
916 CCGGCTGTCTGCCTTACCGGTCTTCTGCGGCGCTGCGGCACCGGGGAC    965

301 GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe  317
    |||||||||||||||||
966 CAAGGCGCGCTCTTTGGAAACCTGTCCACGGCCAACCCGCGGCAGCAGAT  1015

317 tAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGlySerSerP  334
    |||||||||||||||||
1016 GAACGGGCTGACCTCGTTCCTGGACGCGTCCACCGTGTATGGCAGCTCCC 1065

334 roAlaLeuGluArgGlnLeuArgAsnTrpThrSerAlaGluGlyLeuLeu  350
    ||||||||||||::::||||||||
1066 CGGCCCTAGAGAGGCAGCTGCGGAACTGGACCAGTGCCGAAGGGCTGCTC 1115

351 ArgValHisGlyArgLeuArgAspSerGlyArgAlaTyrLeuProPheVa  367
    |||||||||::::||||||||
1116 CGGCTCCACGCGGCGCCTCCGGGACTCCGGCGCGGCCTACCTGCCCTTCGT 1165

367 lProProArgAlaProAlaCysAlaProGluProGlyAsnProGlyS     384
    |||||||||||||||||::::||||||||| ||||||||
1166 GCCCCCACGCGGCGCCTTCGGCCTGTGCGCCCGAGCCCGGCCATCCCGGAG 1215

384 luThrArgGlyProCysPheLeuAlaGlyAspAlaSerGlyArgAlaSerGluVal 400
    ||||||||||||||||||||||:::::|||||
1216 AGACCCGCGGGCCCTGCTTCCTGGCCGGAGACGCGCCGCCACCGAGGTC   1265

401 ProSerLeuThrAlaLeuHisThrLeuTrpLeuTrpLeuArgGluHisAsnArgLe 417
    |||||||||||||||||||||
1266 CCCTCCCTGACGGCACTGCACACGTGTGGCTGCGGCGCACAACCGCCT    1315

417 uAlaAlaAlaLeuLeuLysAlaLeuAsnAlaHisTrpSerAlaAspAlaValT 434
    |||||||||||||||||||||
1316 GGCCGCGGCGCTCAAGGCCCTCAATGCGCACTGGAGCGCGGACGCCGTGT  1365

434 yrGlnGluAlaArgLysValValGlyAlaLeuHisGlnIleIleThrLeu  450
    ||||||||||||||||||
1366 ACCAGGAGGCGCGCAAGGTCGTGGGCGCTGCACTGCACCAGATCATCACCCTG 1415

451 ArgAspTyrIleProArgIleProLeuGlyProGlnAlaPheGlnGlnTyrVa 467
    |||||||||||||||
1416 AGGGATTACATCCCCAGGATCCCTGGACCCCAGGCCTTCCAGCAGTACGT 1465

467 lGlyProTyrGluGlyTyrAspSerThrAlaAsnProThrValSerAsnV  484
    |||||||||||||||||||
1466 GGGTCCCTATGAGGGCTATGACTCCACCGCCAACCCCACTGTGTCCAACG  1515

484 alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu  500
    ||||||||||||||||||
1516 TGTTCTCCACAGCGCCTTCCGCTTCGGCCATGCCACGATCCACCCGCTG   1565

501 ValArgArgLeuAspAlaSerPheGlnGluHisProAspLeuProGlyLe  517
    |||||||||||||||||||
1566 GTGAGGAGGCTGGACGCGAGCTTCCAGGAGCACCCGGACCTGCCCGGGCT  1615

517 uTrpLeuHisGlnAlaPhePheSerProTrpThrLeuLeuArgGlyGlyG  534
    |||||||||||||||||||
1616 GTGGCTGCACCAGGCTTTCTTCAGCCCATGGACATTACTCCGTGGA.....1661

534 lyLeuAspProLeuIleLeuArgGlyLeuLeuAlaArgProAlaLysLeuGln 550
    ||||||||||||||||||
1661 ..................................................1661

551 ValGlnAspGlnLeuMetAsnGluGluLeuThrGluArgLeuPheValLe  567
    ..................................................
1661 ..................................................1661

567 uSerAsnSerSerThrLeuAspLeuAlaSerIleAsnLeuGlnArgGlyA  584
    ..................................................
1661 ..................................................1661

584 rgAspHisGlyLeuLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeu 600
    |||||||||||||||||||||||||
1662 ..........GGTTACAATGAGTGGAGGGAGTTCTGCGGCCTG        1694

601 ProArgLeuGluThrProAlaAspLeuSerThrLeuAlaIleAlaSerArgSe 617
    |||||||||||||||||||||
1695 CCTCGCCTGGAGACCCCCGCTGACCTGAGCACACAGCCCATCGCCAGGAGG 1744

617 rValAlaAspLysIleLeuAspLeuTyrLysHisProAspAsnIleAspV  634
    |||||||||||||||||
1745 CGTGGCCGACAAGATCCTGGACTTGTACAAGCATCCTGACAACATCGATG  1794

634 alTrpLeuGlyGlyLeuAlaGluAsnPheLeuProGlnAlaArgThrGly  650
    |||||||||||||||||
1795 TCTGGCTGGGAGGCTTAGCTGAAAACTTCCTCCCCAGGCTCGGACAGGG   1844

651 ProLeuPheAlaCysLeuLeuGlyLysGlnMetLysGlnLeuArgAspPheVa 667
    |||||||||||||||||||
1845 CCCCTGTTTGCCTGTCTCATTGGGAAGCAGATGAAGCGCTTCGCGGACGG  1894
```

```
mRNA            <1..2877
                /note="TPO mRNA"
gene            73..2703
                /gene="TPO"
CDS             73..2703
                /gene="TPO"
                /note="thyroid peroxidase"
                /codon_start=1
                /db_xref="GDB:G00-120-446"
                /protein_id="AAA61216.1"
                /db_xref="GI:339871"
                /translation="MRALAVLSVTLVMACTEAFPFFISRGKELLWGKPEESRVSSVLE
                ESKRLVDTAMYATMQRNLKKRGILSPAQLLSFSKLPEPTSGVIARAAEIMETSIQAMK
                RKVNLKTQQSQHPTDALSEDLLSIIANMSGCLPYMLPPKCPNTCLANKYRPITGACNN
                RDHPRWGASNTALARWLPPVYEDGFSQPRGWNPGFLYNGFPLPPVREVTRHVIQVSNE
                VVTDDRYSDLLMAWGQYIDHDIAFTPQSTSKAAFGGADCQMTCENQNPCFIQLPE
                EARPAAGTACLPFYRSSAACGTGDQGALFGNLSTANPRQQMNGLTSFLDASTVYGSSP
                ALERQLRNWTSAEGLLRVHARLRDSGRAYLPFVPPRPAACAPEPGIPGETRGPCFLA
                GDGRASEVPSLTALHTLWLREHNRLAAALKALNAHWSADAVYQEARKVVGALHQIITL
                RDYIPRILGPEAFQQYVGPYEGYDSTANPTVSNVFSTAAFRFGHATIHPLVRRLDASF
                QEHPDLPGLWLHQAFFSPWTLLRGGYNEWREFCGLPRLETPADLSTAIASRSVADKIL
                DLYKHPDNIDVWLGGLAENFLPRARTGPLFACLIGKQMKALRDGDWFWWENSHVFTDA
                QRRELEKHSLSRVICDNTGLTRVPMDAFQVGKFPEDFSSCDSIPGMNLEAWRETFPQD
                DKCGFPESVENGDFVHCEESGRRVLVYSCRHGYELGQREQLTCTQEGWDFQPPLCKDV
                NECADGAHPPCHASARCRNTKGGFQCLCADPYELGDDGRTCVDSGRLPRATWISMSLA
                ALLIGGFAGLTSTVICRWTRTGTKSTLPISETGGTPELRCGKHQAVGTSPQRAAAQD
                SEQESAGMEGRDTHRLPRAL"

BASE COUNT      632 a      883 c      829 g      533 t
ORIGIN      Chromosome 2pter-q11

alignment_scores:
       Quality:  4206.50                  Length:     848
       Ratio:    5.345                     Gaps:         1
Percent Similarity:  92.807     Percent Identity:  92.453

alignment_block:
US-08-482-402A-3_COPY_1_848 x HUMTPOB

Align seg 1/1   to: HUMTPOB    from: 1  to: 2877

   1 MetArgAlaLeuAlaValLeuSerValThrLeuValMetAlaCysThrGl  17
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  73 ATGAGAGCGCTCGCTGCGCTGTCTGTCACGCTGGTTATGGCCTGCACAGA 122

  17 uAlaPhePhePheIleSerArgGlyLysGluLeuLeuTrpGlyLysLysP  34
     ||||||||||||||||||||||||||||||||||||||||||||||||
 123 AGCCTTCTTCCCCTTCATCTCGAGAGGGAAAGAACTCCTTGGGGAAAGC 172

  34 roGluGluSerArgValSerSerValLeuGluSerLysArgLeuVal  50
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 173 CTGAGGAGTCGCGTGTCTTCTAGCGTCTTGGAGAAGAACAAGCGCCTGGTG 222

  51 AspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArgGlyIl  67
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 223 GACACCGCCATGTACGCCACGATGCAGAGAAACCTCAAGAAAAGAGGAAT 272

  67 eLeuSerGlyAlaGlnLeuLeuSerPheSerLysLeuProGluProThrS  84
     |||||||||||||||||||||||||||||||||||||||||||||||
 273 CCTTTCTCCAGCTTCGTCGTCTTTTCCAAACTTCCTGAGCCCAACAA 322

  84 erGlyValIleAlaArgAlaAlaGluIleMetGluThrSerIleGlnAla 100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 323 GCCGGAGTGATTGCCCGAGCAGCAGAGATAATGGAAACATCAATACAAGCG 372

 101 MetLysArgLysValAsnLeuLysThrGlnGlnSerGlnHisProThrAs 117
     |||||||||||||||||||||||||||||||||||||||||||||||
 373 ATGAAAAGAAAAGTCAACCTGAAAACTCAAACAGCACATCCAACGGA 422

 117 pAlaLeuSerGluAspLeuLeuSerIleIleAlaAsnMetSerGlyCysL 134
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 423 TGCTTTATCAGAAGATCTGCTGAGCATCATTGCAAACATGTCTGGATGTC 472
```

```
 667 yAspTrpPheTrpGluAsnSerHisValPheThrAspAlaGlnArgA 684
     |||||||||||||||||||||||||||||||||||||||||||||||
1895 TGACTGGTTTGGTGGGAGAACAGCCACGTCTTCACGGATGCACAGAGGC 1944

 684 rgGluLeuGluLysHisSerLeuSerArgValIleCysAspAsnThrGly 700
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1945 GTGAGCTGGAGAAGCACTCCCTGTCTCGGGTCATCTGTGACAACACTGGC 1994

 701 LeuThrArgValProMetAspAlaPheGlnValGlyLysPheProGluAs 717
     |||||||||||||||||||||||||||||||||||||||||||||||
1995 CTCACCAGGGTGCCCATGGACGCCTTCCAAGTCGGCAAATTCCCCGAAGA 2044

 717 pPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgG 734
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2045 CTTTGAGTCTTGTGACAGCATCACTGGCATGAACCTGGAGGCCTGGAGGG 2094

 734 luThrPheProGlnAspAspLysCysGlyPheProGluSerValGluAsn 750
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2095 AAACCTTTCCTCAAGACGACAAGTGTGGCTTCCCAGAGAGCGTGGAGAAT 2144

 751 GlyAspPheValHisCysGluGluSerGlyArgArgValLeuValTyrSe 767
     |||||||||||||||||||||||||||||||||||||||||||||||
2145 GGGGACTTTGTGCACTGTGAGGAGTCTGGAGGGCGTGCTGGTGTATTC 2194

 767 rCysArgHisGlyTyrGluLeuGlyGlnGlyGlnLeuThrCysThrG 784
     ||||||||||||||||||||||||||||||||||||||||||||||||
2195 CTGCCGCCACGGTATGAGCTCCAAGGCCGGAGCAGCTCACTTGCACCC 2244

 784 lnGluGlyTrpAspPheGlnProLeuCysLysAspValAsnGluCys 800
     ||||||||||||||||||||||||||||||||||||||||||||||
2245 AGGAAGGATGGGATTTCCAGCCTCCCCCTCCAAAGATGTGAACGAGTGT 2294

 801 AlaAspGlyAlaHisProProCysHisAlaSerAlaArgCysArgAsnTh 817
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2295 GCAGACGGTGCCCACCCCCCGTGCCACGCCTCTGCGAGGTGCAGAAACAC 2344

 817 rLysGlyGlyPheGlnCysLeuCysAlaAspProTyrGluLeuGlyAspA 834
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2345 CAAAGGCGGCTTCCAGTGTCTCTGCGCGGACCCCTACGAGTTAGGAGACG 2394

 834 spGlyArgThrCysValAspSerGlyArgLeuProArgValThr 848
     ||||||||||||||||||||||||||||||||||||||||||||
2395 ATGGGAGAACCTGCGTAGACTCCGGGAGGCTCCCTCGGGTGACT 2438
```

seq_name: gb_pr:HUMTPOB

```
seq_documentation_block:
LOCUS       HUMTPOB        2877 bp     mRNA            PRI       14-JAN-1995
DEFINITION  Human thyroid peroxidase mRNA, clone phTPO-2.4.
ACCESSION   J02970
VERSION     J02970.1  GI:339870
KEYWORDS    alternative splicing; thyroid peroxidase.
SOURCE      Human thyroid gland (from patients with Grave's disease), cDNA to
            mRNA, clone phTPO-2.4.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2877)
AUTHORS     Kimura,S., Kotani,T., McBride,O.W., Umeki,K., Hirai,K., Nakayama,T.
            and Ohtaki.S.
TITLE       Human thyroid peroxidase: complete cDNA and protein sequence,
            chromosome mapping, and identification of two alternately spliced
            mRNAs
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 84 (16), 5555-5559 (1987)
MEDLINE     87289643
COMMENT     Computer-readable copy of sequence in [1] kindly provided by
            S.Kimura, 09-SEP-1987.
FEATURES             Location/Qualifiers
source          1..2877
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /map="2pter-p24"
```

us-08-482-402a-3_copy_1_848.rge

```
134  euProTyrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLys  150
     |||||||||||||||||||||||||||||||||||||||||||||||||
473  TCCTTACAGTCTGCCCCAAAATGCCCCAAACACTTGCTGGCAACAAA     522

151  TyrArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpG1  167
     |||||||||||||||  |||||||
523  TACAGGCCCATCACCAGGAGCTTGCAACAACAGAGACCACCCCAGATGGGG 572

167  yAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspG  184
     |||||||||||||||||||||||||||||||||||||
573  CGCCTCCAACACGGCCCTGGCACGATGGCTCCCTCCAGTCTATGAGGACG  622

184  lyPheSerGlnProArgGlyTrpAsnProGlyPheLeuTyrAsnGlyPhe  200
     |||||||||||||||||||||||||||||||| |||||
623  GCTTCAGTCAGCCCCGAGGCTGGAACCCCGGCTTCTTGTACAACGGGTTC  672

201  ProLeuProProValArgGluValThrArgHisArgValIleGlnValSerAs  217
     ||||||||||||||||||||||||||||||||||||||
673  CCACTGCCCCCGGTCCGGGAGGTGACAAGACATGTCATTCAAGTTTCAAA  722

217  nGluValThrArgAspAspAspArgTyrSerAspLeuLeuMetAlaTrpG  234
     |||||||||||||||||||||||||||||||||||||
723  TGAGGTTGTCACAGATGATGACCCGTATTCTGACCTCCTGATGGCATGGG  772

234  lyGlnTyrIleAspHisAspIleAlaPheThrProGlnSerThrSerLys  250
     |||||||||||||||||||||||||||||||||||
773  GACAATACATCGACCACGACATCGCCTTCACACCACCAGAGCACCAGCAAA 822

251  AlaAlaPheGlyGlySerAspCysGlnMetThrCysGluAsnGlnAsnAs  267
     |||||||||||||||||||:::|||||||||||||||||||||
823  GCTGCCTTCGGGGGAGGGGCTGACTGCCAGATGACTTGTGAGAACCAAAA  872

267  nProCysPheProIleGlnLeuProGluGluAlaArgProAlaAlaGlyT  284
     |||||||||||||||||||||||||||||||||||||
873  CCCATGTTTTTCCCATACAACTCCCGGAGGAGGCCCGGCCGCGGGGCA    922

284  hrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyThrGlyAsp  300
     ||||||||||||||||||||||||||||||||||||
923  CCGGTGTGCCCTTCTACCGGTCTTCGGCCGCCTGCGGCCACCGGGGGAC   972

301  GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe  317
     |||||||||||||||||||||||||||||||||||||
973  CAAGGCGCGCTCTTTGGGAACCTGTCCACGGCCAACCCCGGGCAGCAGAT 1022

317  tAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGlySerSerP  334
     |||||||||||||||||||||||||||||||||||||
1023 GAACGGGTTGACCTCGTTCCTGGACGCGTCCACCGTGTATGGCAGCTCCC 1072

334  roAlaLeuGluArgGlnLeuArgAsnTrpThrSerAlaGluGlyLeuLeu  350
     |||||||||||||||||||||||||||||||||||||
1073 CGGCCCTAGAGAGGCAACTGCGGAACTGGACCAGTGCCGAAGGGCTGCTC 1122

351  ArgValHisGlyArgLeuArgAspSerGlyArgAlaTyrLeuProPheVa  367
     ||||||||||||||:::|||||||||||||||||||||
1123 CGCGTCCACGGCGCCCTCCGGGACTCCGGCCGCGCCTACCTGCCCTTCGT 1172

367  lProProArgAlaProAlaAlaCysAlaProAlaProGluProGlyAsnProGlyG 384
     ||||||  |||||||||||||||||||||||||||||||
1173 GCCGCCACGCGCGCCTGCGGCCTGCGCGCCCGCTGCCCCGGAGCCCGGAG 1222

384  luThrArgGlyPheCysPheLeuLeuAlaGlyAspAspGlyAlaSerGluVal  400
     ||||||||||||||||||||||||||||||||||||||||||
1223 AGACCCGCGGGGCCTGCGGCCTTCCCTGGAACGGCCCCAGCGGAGGTC   1272

401  ProSerLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLe  417
     ||||||||||||||:::|||||||||||||||||||||
1273 CCCTCCCACGGCACTGCACACCCTGTGGCTGCGCGAGCACAACCGCCT   1322

417  uAlaAlaAlaLeuLysLeuAsnAlaHisTrpSerAlaAspAlaValT      434
     |||||||||||||||||||||||||||||||||
1323 GGCCGCGGCGCTCAAGGCCTGAACGCCCACTGGAGCGCGGACGCCGTGT  1372

434  yrGlnGluAlaArgLysValValGlyAlaLeuHisGlnIleIleThrLeu  450
```

```
     |||||||||||||||||||||||||||||||||||||||||||||||||
1373 ACCAGGAGCGCCGCCAAGGTCGTGGGCGCGCTCTCGCCAGATCATCACCCTG 1422

451  ArgAspTyrIleProArgIleLeuGlyProGluAlaPheGlnIleTyrVa  467
     |||||||||||||||||||||||||||||||||||||||||||
1423 AGGGATTACATCCCCAGGATCCTGGGACCCGAGGCCTTCCAGCAGTACGT 1472

467  lGlyProTyrGluGlyTyrTyrAspSerProThrValSerAsnV        484
     |||||||||||||||||||||||||||||||||||||
1473 GGGTCCCTATGAAGGCTATGACTCCCCCACCGTGTCCAACG          1522

484  alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu  500
     |||||||||||||||||||||||||||||||||||||||||
1523 TGTTCTCCACAGCCGCCTTCCGCTTCGGCCACGCGACCATCCACCCGCTG 1572

501  ValArgArgLeuAspSerPheGlnHisGluHisProAspProGlyLe    517
     ||||||||||||||||||||||||||||||||||||
1573 GTGAGGAGGCTGGACGCCAGCTTCCAGGAGCACCCCGACCTGCCCGGGCT 1622

517  uTrpLeuHisGlnAlaPheSerProThrTrpLeuAlaArgGlyGlyG    534
     ||||||||||||||||||||||||||||||
1623 GTGGCTGCACCAGGCTTCTTCAGCCCATGGACATTACTCCGTGGA.....  1668

534  lyLeuAspProLeuArgGlyLeuLeuAlaArgProAlaLysLeuGln    550
     ||||||||||||||||||||||||||||||
1668 ..................                                  1668

551  ValGlnAspGlnLeuMetAsnGluGlnGluLeuThrGluArgLeuValLe  567
     ||||||||||||||||||
1668 .................                                   1668

567  uSerAsnSerThrLeuAspLeuAlaSerIleAsnLeuGlnArgGlyA    584
     ||||||
1668 .........                                           1668

584  rgAspHisGlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeu  600
     |||||||||||||||
1669              GGTTACAATGAGTGGAGGGAGTTCTGCGGCCTG    1701

601  ProArgLeuGluThrProAlaAspProLeuSerThrAlaIleAlaSerArgSe  617
     ||||||||||||||||||||||||||||||||||||
1702 CCTCGCCTGGAGACCCCCGCTGACCTGAGCACCGCCATCGCCAGCAGGAG 1751

617  rValAlaAspLysIleLeuAspLeuTyrLysHisProAspAsnIleAspV  634
     |||||||||||||||||||||||||||
1752 CGTGGCCGACAAGATCCTGGACCTTGTACAGCATCCTGACAACATCGATG 1801

634  alTrpLeuGlyGlyLeuAlaGluAsnPheLeuProArgAlaArgThrGly  650
     |||||||||||||||||||||||||||||||
1802 TCTGGCTGGGAGGGCTTAGCTGAAAACTTCCTCCCCAGGGCTCGGACAGGG 1851

650  ProLeuPheAlaCysLeuLeuGlyLysGlnMetLysAlaLeuArgAspGl  667
     ||||||||||||||||||||||||||||||||||||||
1852 CCCCTGTTTGCCTGTCTCTCATTGGGAAGCAGAATGAAGGCTCTGCGGATGG 1901

667  yAspTrpPheTrpTrpGluAsnSerHisValPheThrGlnAlaGlnArgA  684
     |||||||||||||||||||||||||||||||||
1902 TGACTGGTTTTGGTGGGAGGAAACAGCCACGTCTTCACGGAATGCACAGAGGC 1951

684  rgGluLeuGlyLysHisSerSerLeuSerArgValIleCysAspAsnThrGly 700
     |||||||||||||||||||||||
1952 GTGAGCTGGAGAAGCACTCCCGTCTCCGGGTCATCTGTGACAACACTGGC 2001

701  LeuThrArgValProMetGluAspAlaPheGlnValGlyLysLysProGluAs 717
     ||||||||||||||||||||||||||||
2002 CTCACCAGGGTGCCCATGGATGCCTTCCAAGTCGGCAAATTCCCTGAAGA 2051

717  pPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgG  734
     ||||||||||||||||||||||
2052 CTTTGAGTCTTGTGACAGCATCCCTGGCATGAACCTGGAGGCCTGGAGGG 2101

734  luThrPheProGlnAspProLysCysGlyPheProGluSerValGluAsn  750
     ||||||||||||||||||||
```

```
2102 AAACCTTCCTCAAGACGACAAGTGTGGCTGGGACTCCGAGAGAGGCGCTGAGAAT 2151
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
 751 GlyAspPheValHisCysGluGluSerGlyArgArgValLeuValTyrSe 767
2152 GGGGACTTTGCACTGTGAGGAGCTCGGAGGCCGCTGGTCTGGTCTATTC 2201

 767 rCysArgHisGlyTyrGluLeuGlyIleGlyArgGluThrCysThrG 784
          ||||||||||||||||||||||||||||||||||||||||||||||
2202 CTGCCGGCACGGGTATGAGCTCGGAAGGCCCGGAGCAGCTCACTTGCACCC 2251

 784 lnGluGlyTrpAspPheGlnProProLeuCysLysAspValAsnGluCys 800
          ||||||||||||||||||||||||||||||||||||||||||||||||||
2252 AGGAAGGATGGGAGTTTCCAGCCTCCCCTGTGCAAAGATGTGAACGAGTGT 2301

 801 AlaAspGlyAlaHisProProCysHisAlaSerAlaArgCysArgAsnTh 817
          ||||||||||||||||||||||||||||||||||||||||||||||||||
2302 GCACAGACGGTGCCCCCCCTGCCACGCCCTCGCCAGGTCCAGAAACAC 2351

 817 rLysGlyGlyPheGlnCysLeuCysSerAlaAspProTyrGluLeuGlyAspA 834
          |||||||||||||||||||||||||||||||||||||||||||||||
2352 CAAAGGCGGCTTCCAGTGTCTCTGCGGCGGACCCCTACGAGTTAGGAGACG 2401

 834 spGlyArgThrCysValAspSerGlyArgLeuProArgValThr 848
          ||||||||||||||||||||||||||||||||||||:||||
2402 ATGGGAGAACCTGCGTAGACTCCGGAGGCTCCCTGCGGCGACT 2445
```

seq_name: gb_ro:RNTPO

```
seq_documentation_block:
LOCUS       RNTPO           3237 bp    mRNA             ROD       12-SEP-1993
DEFINITION  Rat mRNA for thyroid peroxidase.
ACCESSION   X17396 M27275
VERSION     X17396.1  GI:57382
KEYWORDS    thyroid peroxidase.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 3237)
  AUTHORS   Rapoport,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-AUG-1989) to the EMBL/GenBank/DDBJ databases
REFERENCE   2  (bases 1 to 3237)
  AUTHORS   Derwahl,M., Seto,P. and Rapoport,B.
  TITLE     Complete nucleotide sequence of the cDNA for thyroid peroxidase in
            FRTL5 rat thyroid cells
  JOURNAL   Nucleic Acids Res. 17 (20), 8380 (1989)
  MEDLINE   90045972
FEATURES             Location/Qualifiers
     source          1..3237
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /cell_line="FRTL5."
     CDS             42..2786
                     /note="thyroid peroxidase (AA 1-914)"
                     /codon_start=1
                     /protein_id="CAA35257.1"
                     /db_xref="GI:57383"
                     /db_xref="SWISS-PROT:P14650"
                     /translation="MRTLGAMAVMLVVMGTAIFLPFLLRSRDIIGGKTMTSHVISVVE
                     REQSQFSTDALSADILATIANLSGCLPFMLPPRCPDTCLANKYRPITGVCNNRDHPRW
                     GASNTALARWLPPVYEDGFSQPRGWNPNFLYHGFPLPPVREVTRHLIQVSNEAVTEDD
                     QYSDFLPVWGQYIDHDIALTPQSTSTAAFWGGVDCQLTCENQNPCFPIQLPSNSSRTT
                     ACLPFYRSSAACGTGDQGALFGNLSAANPRQQMNGLTSFLDASTVYGSSPGVEKQLRN
                     WSSSAGLLRVNTLHLDSGRAYLPFASAACAPEPGAPHANRTPCFLAGDGRASEVPALA
                     AVHTLWLREHNRLATAFKAINTHWSANTAYQEARKVVGALHQIITMRDYIPKILGPDA
                     FRQYVGPYEGYNTPVNPTVSNVFSTAAFRFGHATVHPLVRRLNTFDQDHTELPRLQLH
                     DVFFRPWRLIQEGGLDPIVRGLLARPAKIQVQEOLMNEELTERLFVLSNVGTLDLASL
                     NLQRGRDHGLPGYNEWREFCGLSRLDTGAELNKAIANRSMVNKIMELKHADNIDVWL
                     GGLAEKFLPGARTGPLFACIIGKQMKALRDGDRFWWENSHVFTDAQRQELEKHSLPRV
                     ICDNTGLIRVPVDAFRIGKFPQDFESCEEIPSMDLRLWRETFPQDDKCVFPEKVDNGN
```

Right column (protein sequence top):
```
FVHCEESGKLVLVYSCPHGYKLQGQEQVTCTQNGWDSEPPVCKDVNECADLTHPPCHS
SAKCKNTKGSFSQCVCTDPYMLGEDEKTCIDSGRLPRASWVSIALGALLIGGLASLSWT
VICRWTHADKKSTLLITRVTMESGFRKSQESGISPQKAEVQDAEPAYGSRVLLCE
"
```

alignment_scores:
```
BASE COUNT  816 a  874 c  831 g  716 t
ORIGIN
```

```
alignment_scores:
                Quality:  3563.00              Length:   848
                Ratio:       4.591              Gaps:       3
Percent Similarity:  91.509    Percent Identity:  76.651
```

```
alignment_block:
US-08-482-402A-3_COPY_1_848 x RNTPO

Align seg 1/1 to: RNTPO    from: 1  to: 3237
```

```
   1 MetArgThrLeuGlyAlaMetAlaValMetLeuValValMetGlyThrAlaIlePheLeu 17
          |||||::|||:||||:|:||||||||||::||   |||||:::||
  42 ATGAGAACACTTGGAGCTATGGCAGTAATGCTGGTTGTGATGGGAACTGC 91

  17 ProPheLeuLeuArgSerArgAspIleIleGlyGlyLysThrMetThrSerHisVallleSerVal 34
          |:::|||:::|||::||||       |||||:||||||||||||::|
  92 AAGTTTCCTCCCTTTCCTCCTCGAGAAGCAGCAGACATCTTGGGGGGAAGA 141

  34 GluGluArgGluGlnSerGlnPheSerThrAspAlaLeuSerAlaAspIleLeuAlaThrIle 50
          ||::|||::|||:::||||||:||||       |||||||:::||
 142 GAACAGTCCCAGTTTTCAACAGATGCTTTGTCTGCAGACATACTAGCCACA 191
```

us-08-482-402a-3_copy_1_848.rge

```
534 lyLeuAspProLeuIleArgGlyLeuLeuAlaAlaArgProAlaAlaLysLeuGln 550
    |||||||||||::::::|||||||||||||||||||||||||||||
1606 GTTTGGATCCAATAGTAAGAGGCCCTCCTGGCAAGACCAACCAGCTGCAA 1655

551 ValGlnAspGlnLeuMetAsnGluGlyLeuLeuThrGluArgLeuPheValLe 567
    ||||||||||||||||||||||||||||||||:::||||||||||
1656 GTACAGGACCAACTGATGAATGAGGAACTGACCGAGAGCCTCTTTGTGTT 1705

567 uSerAsnSerThrLeuAspLeuAlaSerIleAsnLeuGlnAsnArgGlyA 584
    |||||||:::||||||||||||||||:::||||||||||||
1706 GTCTAATGTGGCACCTTGGATCTGCATCACTGAACTTGCAGAGGGGCC 1755

584 rgAspHisGlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeu 600
    |||||||||||||||||||||||||||||||||||||||||
1756 GGGGACCACGGCTTACCAGGCTACAATGAGGTGGAGAGAGTTCTGTGGCTTG 1805

601 ProArgLeuGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSe 617
    ||||||||   :::||||||||||||||||||||
1806 TCACCCCTGGACACAGGAGCTGAGCTGAACAGGCCATTGCCAACAGAAG 1855

617 rValAlaAspLysIleLeuAspLeuTyrLysHisProAspAsnIleAspV 634
    |:::||||||   :::|||||||||||||||||||||
1856 CATGTCACCAAGATAATGAGTATACAGCATGTGACAACATTGATG 1905

634 alTrpLeuGlyGlyLeuAlaGluAsnPheLeuProArgAlaArgThrGly 650
    ||||||||||||::::::||||||  ||||||:::||||||||||
1906 TCTGGTTGGGAGGCCTGGCTGAAAAGTTCTTACCGGGAGCCCGCACCGGT 1955

651 ProLeuPheAlaCysLeuGlyLeuGlyLysGlnMetLysAlaLeuGluArgAspGl 667
    ||||||||||||||||||||||||||||||||||||
1956 CCCTCGTTTGCATGTATCATTGGGAGCAGAGTGAAGGCTCGAGGGGATGG 2005

667 yAspTrpPheTrpTrpGluAsnSerHisValPheThrAspAlaGlnArgA 684
    |:::||||||||||||   |||||||||||||||||:::
2006 GGACAGGTTTTGGTGGGAACAGCCATGTCTTCACAGACGCCTCAGAGGC 2055

684 rgGluLeuGluLysHisSerLeuSerArgValIleCysAspAsnThrGly 700
    :::||||||||||||   |||||||||||||||||||||
2056 AGGAACTGGAAAAGCATTCACTACCTCGGTCATCTGTGACAACACCGGC 2105

701 LeuThrArgValProMetAspAlaPheGlnInValGlyLysPheProGluAs 717
    ||||||||||||   :::|||||||||||||||||||
2106 CTCACCAGAGTACCTGTGGATGCCTTCCGTATTGGAAAGTTCCCCGGA 2155

717 pPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgG 734
    :::||||||||||||||||||||||||||||||||||
2156 CTTTGAATCCTGTGAGAAATCCCTAGCATGGAACCTCAGAGCTGTGGAGGG 2205

734 luThrPheProGlnAspLysCysGlyPhePheProGluSerValGluAsn 750
    |||||||||||||||||||::::::||||||||||||||||
2206 AGACCTTCCCCAAGACAAGTGTCTTCCCAGAAGAAGGTGGACAAT 2255

751 GlyAspPheValHisCysGluGlySerGlyArgArgValLeuValTyrSe 767
    ||||||||||::::::||||||||||||||||||||||
2256 GGGGACTTTGTGCACTGTGAAGAATCTGGGAGGCGTGTACTGGTGTATTC 2305

767 rCysArgHisGlyTyrGluGlnGlyGlyLeuGlnLeuThrCysThrG 784
    ||||||||||||::::::|||||||||||||
2306 CTGTTTCCATGGATACAAGCTGCAAGGCCAGGAGCAGAGCCACATGTACCC 2355

784 lnGluGlyTrpAspAspPheGlnProLeuCysLysAspValAsnGluCys 800
    |||||::::::||||||||||||||||||||
2356 AGAATGGATGGGACTCAGAGCCCTGTCTGTAAAGATGTTAATGAGTGT 2405

801 AlaAspGlyAlaHisProProCysHisAlaSerAlaArgCysArgAsnTh 817
    |||::::::|||||||||||||||||
2406 GCAGATCGCACACACCCTTGCCACTCCTGCCAAAGTGCAAGACACAC 2455

817 rLysGlyGlyTyrGlnCysLeuCysAlaAspProTyrGluLeuGlyAspA 834
    |:::||||||||||||||||||||
2456 CAAGGAAGCTTCCAGTGTGTGCACAGACCCCTACATGCTAGGTGAGG 2505

834 spGlyArgTyrCysValAspSerGlyArgLeuProArgValThr 848
```

```
234 lyGlnTyrIleAspHisAspIleAlaPheThrProGlnSerThrThrLys 250
    |||||||||||||||||||||||||||||||||||||||||||
724 GACAGTACATCGATCATGACATTGCTCTCACACCAGAGCACCAGCACA 773

251 AlaAlaPheGlyGlySerAspCysSerGlnMetThrCysGluAsnGlnAs 267
    ||||||   |||||||||||||   |||||||||||||||||
774 GCAGCCTTCTGGGGAGGTGTCGACTTGACCTGTGAGAACCAAAA 823

267 nProCysPheProIleGlnLeuProGluGlnAlaAlaArgProAlaAlaGlyT 284
    ||||||||||||||||||::::::   :::
824 CCCCTGCTTCCCATACAGCTTCCCTCAAACTCC......TCACGGACCA 867

284 hrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyThrGlyAsp 300
    ||||||||||||||||||||||||||||||:::::||||
868 CTGCAATGCCTACCTTTCTACCGCTCCAGCCTCCAGCGCCCGCTGGGGAC 917

301 GlnGlyAlaLeuPheGlyAlaLeuAsnLeuSerThrAlaAsnProArgGlnIleMe 317
    ||||||||||||||||||||||||||||||||||||||||
918 CAAGGACCACTCTTTGGCAACCTGTCTCAGCCAATCCGAGGCAGCAGAT 967

317 tAsnGlyLeuThrSerPheLeuAlaSerThrValTyrGlySerSerP 334
    ||||||||||||||||||||||||||||||||||||||||
968 GAATGGCTTGACCTCCTTCCTTCCTTGAATGGCTTCCACGGTATGGCAGCTCCC 1017

334 roAlaLeuGluArgGlnLeuArgGlnAsnTrpThrSerArgAlaGluGlyLeuLeu 350
    ||||||||||||::::|||||   ||||||||||||||||
1018 CAGGTGTTGAGAAGCAGTTGCGCAACTGGACCAGCGCTCTGCAGGGCTG 1067

351 ArgValHisGlyAlaLeuArgArgAspSerGlyArgAlaTyrArgLeuProPheVa 367
    ||||||:::||
1068 CGTGTCAACACCCTCCACCTAGACTCTGGCCGTGCCTACCGTGCCTTC.. 1115

367 lProProArgAlaProAlaAlaCysAlaProGlyProGlyAsnProGlyG 384
    |||                 ||
1116 .......GCATCAGCGCCCGCTGTGTCCAGAGCCTGGTGCCTCCACACG 1155

384 luThrArgGlyTyrProCysPheLeuAlaGlyAlaGlyAspAlaSerGlyLuVal 400
    |||||||||||||||||||||||:::
1156 CCAACCGCACGCCCCGCTTCCTGGCTGGTGATGGCCGCGCCAGCGAGGTC 1205

401 ProSerLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLe 417
    |||::::::||||||||||||||||||   :::||||||
1206 CCCGCCCTGGCAGTGCACACCTTGTGGCTGCGCGAGCACAACCGCCT 1255

417 uAlaAlaLeuAlaLysAlaLeuAsnAlaHisTrpSerAlaAspAlaValT 434
    ||||::::::||||||   ||||::||||
1256 GGCTACAGCTTTCAAAGCCATCAACACCACATTGGAGCGCCAACACTGCCT 1305

434 yrGlnGluAlaArgLysValValGlyAlaLeuHisGlnIleIleGluThrLeu 450
    ||||||||::::||||||   :::||||||
1306 ACCAGGAGGCACGCAAGGTGGTGGGAGCATTGCACCAGATCATCACCACTG 1355

451 ArgAspTyrIleProArgIleLeuGlyProGluAlaPheGlnGlnTyrVa 467
    ||||::::::|||||||||||||   ::||||||
1356 AGGGATTATATCCCCAGAATCCTGGGTCCGGAGGCCTTCAGGCAGTATGT 1405

467 lGlyProTyrGlyTyrAspGlyTyrAspGlyPheThrProSerValSerAsnV 484
    ||||||||||  |||||::::
1406 GGGGCCCTATGAAGGCTATAACCCCAGTGAACCCTACTGTGTCCAACG 1455

484 alPheSerThrAlaAlaPheArgPheGlyHisAlaLeuHisIleIleHisProLeu 500
    ||||||   ||||||||||||:::
1456 TCTTCTCTACTGCTGCCTTCCGCTTTGGCCATGCCCACCAGATCCTCACTCTG 1505

501 ValArgAspAspProGlnSerPheGlnGluHisGluGlnProProGlyLyGly 517
    ||||||::||||||   :::||||||
1506 GTGAGAGACGACTAAACACCGGACTTCCAGGAGCACGAGCAGCCCCCAGGCT 1555

517 uTrpLeuHisGlnIleAlaPhePheSerProTrpThrLeuLeuArgGlyGlyGlyG 534
    ||||||   |||   :::||||||
1556 GCAGCTGCACGACGTCTTCTTCAGGCCCTGGAGCTTATCCGAGGAGGTG 1605
```

```
                ||  ::||||||||||||||||||||||||||||||||::::
       2506 ATGAGAAGACCTGCATAGATTCTGGCAGGCTACCTCGGGCATCC 2549

seq_name: gb_ro:MMTPER

seq_documentation_block:
LOCUS       MMTPER        3291 bp    mRNA           ROD        25-AUG-1999
DEFINITION  Mus musculus mRNA for thyroid peroxidase.
ACCESSION   X60703
VERSION     X60703.1  GI:297160
KEYWORDS    peroxidase.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3291)
  AUTHORS   Kotani,T., Umeki,K., Yamamoto,I., Takeuchi,M., Takechi,S.,
            Nakayama,T. and Ohtaki,S.
  TITLE     Nucleotide sequence of the cDNA encoding mouse thyroid peroxidase
  JOURNAL   Gene 123 (2), 289-290 (1993)
  MEDLINE   93154601
REFERENCE   2  (bases 1 to 3291)
  AUTHORS   Sachiya,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-1991) O. Sachiya, Central Laboratopy for Clinical
            Investigation, Medical College Hospital, Miyazaki Medical College,
            Kiyotake, Miyazaki 889-16, JAPAN
FEATURES             Location/Qualifiers
     source          1..3291
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /tissue_type="thyroid"
                     /cell_type="follicular"
                     /clone="pmTPO-75"
     mRNA            74..3281
                     /partial
                     /note="peroxidase"
                     /evidence=experimental
     CDS             74..2818
                     /EC_number="1.11.1.7"
                     /codon_start=1
                     /product="peroxidase"
                     /protein_id="CAA43114.1"
                     /db_xref="GI:4539541"
                     /db_xref="MGD:MGI:98813"
                     /db_xref="SWISS-PROT:P35419"
                     /translation="MRTLGAMAIMLVVMGTVVIFLSFILRSRDILCGKTMKSHVISAVE
                     TSQLMVDHAVYNTMKRNLKKREVLSPAQLLSFFKLPESTSGAISRAAEIMETSIQVMK
                     REQSFSTDALSADILGTIANLSGCLPFMLPPRCPDTCLANKYRPITGACNNRDHPRW
                     GASNTALARWLPPVYEDGFSQPKGWNPNFLYHGFPLPVYREVTRHLIQVSNEAVTEDD
                     QYSDFLPVWGQYIDHDIAITPQSTSTAAFWGGVDCQLTCENQNPCFPIQLPSNSSGTT
                     ACLPFYRSSAACGTGDQGALFGNLSAANPRQQMNGLTSFLDASTVYGSSPGVEKQLRN
                     WSSSAGLLRVNTHLHLDAGRAYLPFATAACAPEPGTPRTNRTPCFLAGDGRASEVPALA
                     AVHTLWLREHNRLASAFKAINKHWSANTAYQEARKVVGALHQIITMRDYIPKILGPDA
                     FRQYVGPYEGYNPTVNPFVSNIFSTAAFRFGHATVHPLVRRLNTDFQEHTELPRIQLR
                     DVFFRPWRLIQEGGLDPIVRGLLARAAAKLQVGQQLMNEELTERLFVLSNVGTLDLASL
                     NLQRGRDHGLPDYNEWREFCGLSRLETPAELNKAIANRSMVNKIMDLYKHADNIDVWL
                     GGLAEKFLPGARTGPLFACIIGKQMKALRDGDRFWWENTNVFTDAQRQELEKHSLPRV
                     ICDNTGLTRVPVDAFRIGKFPQDDFESCEDIPSMDLELWRETFPQDDKCVFPEEVDNGN
                     FVHCEESGKLVLVYSCFHGYKLQQEQVTCTQKGWDSEPPVCKDVNECADLTHPPCHP
                     SAQCKNTKGSFQCVCTDPVLGEDEKTCIDSGRLPRASWVSIALGALLIGGLASLTWI
                     VICRWTHADKKATLPITERVTTQSGCRKSQGRGISPHKAAAQDTGQEPASGSRVLLCE
                     "

BASE COUNT    819 a    897 c    854 g    721 t
ORIGIN

alignment_scores:
       Quality: 3520.00          Length:  848
         Ratio: 4.565              Gaps:    3
Percent Similarity: 90.920    Percent Identity: 75.825
```

```
alignment_block:
US-08-482-402A-3_COPY_1_848 x MMTPER                 ..

Align seg 1/1  to: MMTPER   from: 1  to: 3291

   1 MetArgArgArgAlaLeuAlaLeuSerValThrLeuValMetAlaCysThrGl  17
     |||||::||||
  74 ATGAGAACACTTGGAGCTATGCAATAATGCTGGTGGTTATGGGAACTGT  123

  17 uAlaPhePheProPheIleSerArgArgGlyLysGluLeuLeuTrpGlyLysP  34
     |||::: |||||| ||:::::|||::|||||
 124 AATTTCCTCTCTTTTATCCTGAGAAGCAGAGACATCTTGTGTGGGAAGA  173

  34 roGluGluSerArgValSerSerValLeuGluSerArgLysArgLeuVal  50
     |::||| ||||| |||::::::||  |||:::  ::|||
 174 CCATGAAGTCCCATGTTATCAGTGCTGTGGAAACAGCCAGCTCATGGTG  223

  51 AspThrAlaMetTyrAlaThrMetGlnArgAsnLeuLysLysArgGlyIl  67
     |||:: |||||||||::||||||||||||||||||  ::|
 224 GACCATGCAGTCTACAACACCATGAAAGAGAAACCTCAAGAAAAAGGGAAGT  273

  67 eLeuSerGlyAlaGlnLeuLeuSerPheSerLysLeuProGluProThrS  84
     ::||||| ||||||||||||||| |||::||||||  ||||
 274 CCTTTCTCCAGCCACGTTCTCTTCTTCTTTAAGCTGCCGAGTCCACCA  323

  84 erGlyValIleAlaArgAlaAlaGluIleMetGluThrSerIleGlnAla  100
     ||||||::||||||||||||||:::|||||||||||||||||||::::::
 324 GTGGGGCTATTTCCCAGCACGCAGAGATTATGGAAACATCAATACAAGTC  373

 101 MetLysArgLysValAsnLeuLysThrGlnSerGlnHisProThrAs  117
     |||||||          ::||||||  ::|||||||::|||
 374 ATGAAACGT.........GAACAGTCACAGTTCTCCACGGA  405

 117 pAlaLeuSerGluAspLeuSerIleIleAlaAsnMetSerGlyCysL  134
     |||||||| |||::||  ||||||::::||:||:||||||
 406 TGCCCTATCAGCAGACATTCTGGGCACAATTGCCAACCTGTCAGGATGCT  455

 134 euProTyrMetLeuProProLysCysProAsnThrCysLeuAlaAsnLys  150
     ::||||| ::||||||||||||||:||||||||||||||||
 456 TGCCTTTCATGCTGCCCAAGATGTCCGACACCTGCCTGGCCAAATAAG  505

 151 TyrArgProIleThrGlyAlaCysAsnAsnArgAspHisProArgTrpGl  167
     |||::|||||||||  ::||||||||||||||||||||||||||||||
 506 TACCGGCCCATCACAGGGGCTGCAACAATAGAGATCACCCCAGATGGGG  555

 167 yAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspG  184
     |||||||||||||||| ||||||||||||||||||||||||||||||||
 556 AGCCTCCAACACCAGCCTAGCAAGATGGCTGCCTCCTGTCTATGAAGATG  605

 184 lyPheSerGlnProArgGlyTrpAsnProGlyPheLeuTyrAsnGlyPhe  200
     ||||::||||||||| |||||||||||||||  ::|||
 606 GCTTCAGTCAGCCCAAGGCTGGAACCCTAATTCTATACCACGGCTTC  655

 201 ProLeuProProValArgGluValThrArgHisValIleGlnValSerAs  217
     ||||||||||||||||||||::::|||||||||||:::||||||||
 656 CCACTGCCCCCGGTACGGGAAGTGACAAGGCACCTCATTCAAGTTTCCAA  705

 217 nGluValThrArgAspAspAspArgTyrSerAspLeuMetAlaTrpPG  234
     ||::::||||||||| |||||||||||::  ::|||
 706 TGAGGCTGTGACCGAAGATGACCAGTACTCTGATTTTCTGCCGGTGTGGG  755

 234 lyGlnTyrIleAspHisAspIleAlaPheThrProGlnSerThrSerLys  250
     |||||||||||||||||||| ||||||||||||||||||||||||||||
 756 GACAGTACATCGATCATGACATTGCTCTCACCACCAGAGCACTAGCACA  805

 251 AlaAlaPheGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAs  267
     ||||||||||||||||||||||||||||||||||||||||||||||||
 806 GCAGCCTTCTGGGAGGTGTCCAGCTGCCAGCTGCACCTGTGAGAACCAAAA  855

 267 nProCysPheProIleGlnLeuProGluGluAlaArgProAlaAlaGlyT  284
     |||||||||||||||||||::::::||| |||::::: ||
 856 TCCTTGCTTCCCCATACAGCTTCCCTCAAACTCC.....TCAGGGACCA  899
```

```
 284 hrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyThrGlyAsp 300          ||||||||||||
     ||||||||||||||||||||||||||||||||||||||||||||||||||          GGGATCATGGCTTACCAGAGACTACAATGAATGAGGAGAGAGTTCTGTGGCTTG 1837
 900 CTGCGATGCCTGCCTTTCTACCGCGTTCTACCGCTGGGGAC 949

 301 GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe 317     601 ProArgLeuGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSe 617
     ||||||||||||||||||||||||||||||||||||:::||:::          ||||||||||||||||||||||||||||:::||:::||
 950 CAAGGTGCTCTCTTGGCAACCTGTCGCAGCCAATCCGAGGCAGCAGAT 999     1838 TCACGCCTGGAGACCACCAGCTGAGCTGACAACAAGGCCATTGCCAACAGAAG 1887

 317 tAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGlySerSerP 334     617 rValAlaAspLysIleLeuAspTyrLysHisProAspAsnIleAspV 634
     ||:::||||||||||||||||||||||||||||||||||||          ||||||||||||||||||||||||||||||||||
1000 GAATGGCCTTGACCTCTTCCTTCCACTGTCTACGGACAGCTCCC 1049     1888 CATGTCAACAGATAATGGACTTATACAAGCATTGCTGACAACATTGACG 1937

 334 roAlaLeuGluArgGlnLeuArgAsnTrpThrSerAlaGluGlyLeuLeu 350     634 aTrpLeuGlyLeuAlaGluAsnPheLeuGluProArgAlaArgThrGly 650
     ||:::::||||||||||||||||||||||||||||          |||||||||||||||||||||||||||||||||
1050 CTGCGCGTTGAGAAGCAGCAGTTGCGCCACTGACCTCGGCAGGACTGCTG 1099     1938 TCTGGTTGGGAGGCTTGGCTGAAAAGTTCTTGCCGGGGCCCGCACTGGT 1987

 351 ArgValHisGlyArgLeuArgAspSerGlyArgAlaTyrLeuProPheVa 367     651 ProLeuPheAlaCysLeuIleGlyLysGlnMetLysAlaLeuArgAspGl 667
     ||||||:::          ||:::||||||||||||||||||||          |||||||||||||||||||||||||||||||||||||||||
1100 CGTGTCAACACTCCCACCTAGATGCTGGCCGTGCCTACCTGCCCTTC.. 1147     1988 CCTCTGTTTGCATGTATCATTGGGAAGCAGATGAAGGCTCTGAGGGATGG 2037

 367 lProProArgAlaProAlaAlaCysAlaProGluProGlyAsnProGlyG 384     667 yAspTrpPheTrpProGluAsnSerHisValPheThrAspAlaGlnArgA 684
     ||          |||||||||||||||||||||||||||||||          ||||||||||||||||||||||||||||||||||||||||||
1148 ..........GCAACAGCCGCCTGCGCTCCAGAGCCTGGTACCCCCACGCA 1187     2038 GGACAGGTTTGGTGGGGAGAACACCAATGTCTTCACAGACGCTCAGAGGC 2087

 384 luThrArgGlyProCysPheLeuAlaGlyValAspSerGlyArgAlaSerGluVal 400     684 rgGluLeuGluLysHisSerLeuSerArgValIleCysAspAsnThrGly 700
     ::||||          ||||||||||||||||||||||||||||||||          :||||||||||||||||||||||||||||||||||||||||
1188 CCAACCCCACGCCCTGCTCCTGGCAACGTCGCCGCCAGTGAGGTC 1237     2088 AGGAACTAGAAAAGCAATTCACTACCTCGGGTCATCTGTGACAACACTGGA 2137

 401 ProSerLeuThrArgLeuHisThrLeuTrpLeuArgGluHisAsnArgLe 417     701 LeuThrArgValProMetAspAlaPheGlnValGlyLysPheProGluAs 717
     ||||||||||||||||||||||||||||||||||||          ||||||||||||||||||||||||||::||:::|||
1238 CCTGCCCTGCAGCGGTACACACCTTGTGGCTGCGGGAGCACACCGCCT 1287     2138 CTCACCAGAGTACCTGTGGATGCCTTCCGTATTGGAAAGTTCCCCCAAGA 2187

 417 uAlaAlaAlaLeuLeuLeuAsnAlaHisTrpSerAlaAspAlaValT 434     717 pPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAlaTrpArgG 734
     ||||||||||||||||||||          ||||          ||||||||||||||||||||||||||||||||
1288 GGCTTCGGCCTTCAAGGCCATTAACAAGCACTGAGCCGCCAACACTGCCT 1337     2188 CTTTGAATCCTGTGAGGACATCCCGAGCATGGAACCTTGACCTATGGAGGG 2237

 434 yrGlnGluAlaArgLysValValGlyAlaValLeuHisGlnIleIleThrLeu 450     734 luThrPheProGlnAspAspLysCysGlyProGluSerValGluAsn 750
     |||||          |||||||||||||||||||||||||||          |||||||||||||||||||||||||||:||||:::|
1338 ACCAGGAGGCGCGCAAGGTGGTAGGGCACTGACCACCAGATCATCACCATG 1387     2238 AGACCTTCCCACAAGATGACAAGTGTGTCTTCCCAGAGGAGGTGGACAAT 2287

 451 ArgAspTyrIleProArgIleLeuGlyProGluAlaPheGlnTyrVa 467     751 GlyAspPheValHisCysGluGluSerGlyTyrArgArgValLeuValTyrSe 767
     ||||||||||||||||||||||          |||||||||||||||::: |||||||||||||
1388 AGGGATTATATCCCCAAAATCCTGGGTCCTGGCCTTCAGGCAGTATGT 1437     2288 GGGAACTTTGTGACACTGTGAAGAGTCTGGGAAGCTGGTCGTTGTATACTC 2337

 467 lGlyProTyrGluGlyGlyTyrAspSerThrAlaAsnProThrValSerAsnV 484     767 rCysArgHisGlyTyrGluLeuGlnGlyArgGluGlnLeuThrCysThrG 784
     |||||||||||||||          |||::::::||||||||||||          ||||||||||||||||||||||||||:::|||
1438 GGGCCCCATGAAGGCTACAACCCCACTGTGAGCAACTGTGTCCAACA 1487     2338 CTGTTTCCATGGATACAAGCTGCAGGGCCGCAGGGAGCAGGTCACATGTACCC 2387

 484 alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu 500     784 lnGluGlyTrpAspPheGlnProProLeuCysLysAspValAsnGluCys 800
     :::||||||||||||||||||||||:::||||||||          ||||||||||||||||||||||||||:::||||:::|
1488 TCTTCTCCACCGCTGCCTTTCGCTTTGGCCATGCCACAGTCCATCCACTG 1537     2388 AGAAGGGATGGGACTCCCAACCCTCCTGTCCTGTCTAAAGATGTTAATGAGTGT 2437

 501 ValArgLeuAspAspSerPheGlnHisProAspLeuProGlyLeu 517     801 AlaAspGlyAlaHisProProCysHisAlaSerAlaAlaArgCysAsnTh 817
     |||||||||||||||||          ::|||||||||||          ::||||||||:::||:::||||
1538 GTGAGACGCGATAGTGAGAGGCCTAAACACTGACTTCCAGGAGCACACAGA 1587     2438 GCAGATCTGACACACCCACCTTGCCACCCTCTGCCACAGTGCCAAGAAACAC 2487

 517 uTrpLeuHisGlnAlaPhePheSerProTrpLeuLeuArgGlyGlyG 534     817 rLysGlyGlyTyrPheGlnLnCysLeuCysAlaAspProTyrGluLeuGlyAspA 834
     ||||||||||||||||||||||||||||||||||||||          ::::||||||:::||||
1588 GCAGCTGCGTGATGTCTTCTTCAGACCCTGGAGGCTTATCCAGGAAGGTG 1637     2488 CAAGGGAAGCTTCCAGTGTGTGCACAGACCCCTATGTCTAGTGAGG 2537

 534 lyLeuAspProLeuIleArgGlyLeuLeuAlaArgProAlaLysLeuGln 550     834 spGlyArgThrCysValAspSerGlyArgLeuProArgValThr 848
     ||||||||||||::::::||||||||||||||||||          ||          |||||||||||||||||||||||||||||||||||
1638 GTTTGGATCCGATAGTGAGACGCCTCCTGGCGAGCAGCCAAGCTGCAA 1687     2538 ATGAGAAGACCTGCATAGATTCTGGCAGGCTACCTCGGGCATCC 2581

 551 ValGlnAspHisLeuMetAsnGluLeuLeuThrGluArgGlyLeuIle 567
     ||||||          |||||||||||||||||||||||||||||||||||
1688 GTCCAAGGGACAGCTGATGAATGAGGAGCTGACCGGAGAGCTCTTCGTGCT 1737

 567 uSerAsnSerThrLeuAspLeuAspAlaSerIleIleAsnLeuGlnArgGlyArgA 584
     |||||          ||||||||||||||||||||||||||||||||||||
1738 GTCTAACGTGGGACCTTGGATCTGACATCACTCGAACCTGCAGAGGGGCC 1787

 584 rgAspHisGlyLeuProGlyTyrAsnGluTrpArgGluPheCysGlyLeu 600
```

```
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
          Rattus.
REFERENCE   1  (bases 1 to 2777)
AUTHORS     Isozaki,O., Kohn,L.D., Kozak,C.A. and Kimura,S.
TITLE       Thyroid peroxidase: Rat cDNA sequence, chromosomal localization in
            mouse, and regulation of gene expression by comparison to
            thyroglobulin in rat FRTL-5 cells
JOURNAL     Mol. Endocrinol. 3, 1681-1692 (1989)
MEDLINE     90114171
FEATURES             Location/Qualifiers
source          1..2777
                /organism="Rattus norvegicus"
                /db_xref="taxon:10116"
CDS             <1..2313
                /note="thyroid peroxidase"
                /codon_start=1
                /protein_id="AAA42265.1"
                /db_xref="GI:207435"
                /translation="YRPITGVCNNRDHPRWGASNTALARWLPPVYEDGFSQPRGWNPN
                FLYHGLGLPSVREVTRHLIQVSNEAVTEDDQYSDFLPVWAQYIDHDIALTPQSTSTAA
                FWGGVDCQLTCENQNPCFPIQLPSNSSRTTACLPFYRSSAACGTGDQALFGNLSAAN
                PRQQMNGLTSFLDASTVYGSSPGVEKQLRNWSSSAGLLRVNTLHLDSGRAYLPFASAA
                CAPEPGAPHANRTPCFLAGDGRASEVPALAAVHTLWLREHNRLATAFKAINTHWSANT
                AYQEARKVVGALHQIITMRDYIPKILGPDAFRQYVGPYEGYNPTVNPTVSNVFSTAAF
                RFGHATVHPLVRRLNTDFQDHHTELPRLQLHDVFFRPWRLIQEGGLDPIVRGLLARPAK
                LQVQEQLMNEELTERLFVLSNVTLDLASLNLQRGRDHGLPGYNEWREFCGLSRLETP
                AELNKAIANRSMVNKIMELYKHADNIDVWLGGLAEKFLPGARTGPLFACIIGKOMKAL
                RDGDRFWENSHVFTDAQRQELEKHSLPRVICDNTGLTRVPVDAFRIGKFPQDFESCE
                EIPSMDLRLWRETFPQDDKCVFPEKVDNGNFVHCEESGKLVLYSCFHGYKLQGQEQV
                TCTQNGWDSEPPVCKDVNECADLTHPCHSSAKCKNTKGSFQCVCTDPYMLGEDEKTC
                IDSGRLPRASWVSIALGALLIGGLASLSWTVICRWTHADKKSTLLITERVTMESGFRK
                SQESGISPQKAEVQDAEOEPAYGSRVLLCE"
BASE COUNT      690 a       752 c       722 g       613 t
ORIGIN
```

```
alignment_scores:
          Quality:  3082.00          Length:         698
          Ratio:      4.734          Gaps:             2
Percent Similarity:  93.266   Percent Identity:  79.226

alignment_block:
US-08-482-402A-3_COPY_1_848 x RATTPOFR

Align seg 1/1   to: RATTPOFR   from: 1   to: 2777
```

```
151 TyrArgProIleThrGlyValCysAsnAsnArgAspHisProArgTrpGl 167
    |||||||||||||||||||::||::|||||||||||||||||||||||||
  1 TACCGGCCCATCACAGGGGTGTGCAACAACAGAGATCACCCCAGATGGGG  50

167 yAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTyrGluAspG 184
    ||||||||||||||||||::||||||||||::|||||||||||||||::|
 51 AGCCTCCAACACAGCCCTAGCCAGATGGCTGCCTCCTGTCTACGAAGATG 100

184 lyPheSerGlnProArgGlyTrpAsnProValTyrLeuTyrAsnGlyPhe 200
    :||||||::|||||||||||::|||||:||||||::||:|||:||:||::
101 GTTTCAGTCAGCCCAGAGGCTGGAACCCTAATTTCTTATACCATGGCTTG 150

201 ProLeuProProValArgGluValThrArgHisValIleGlnValSerAs 217
    |||||:||||::|||||||:|::|||||:||||||||:||::|||||||||
151 GGACTCCGCTCGGTACGGGAGGTGACAAGACACCTCATTCAAGTTTCAAA 200

217 nGluValValThrArgAspAspArgTyrSerAspLeuLeuMetAlaTrpG 234
    ||::||||::||:::||::||||||::|||||:||||||||:::|||:|||
201 TGAGGCTGTGACTGAAGATGACCTACTCTGATTTCTGCCGGTGTGGG 250

234 lyGlnTyrIleAspHisAspIleAlaPheThrProGlnSerThrSerLys 250
    ::||:||||:|||||||::|||||||||||::||||||||||||||||||
251 CACAGTACATCATGCATGACATTGCTCTCACCACACAGAGCACCAGCACA 300

251 AlaAlaPheGlyGlyGlySerAspCysGlnMetThrCysGluAsnGlnAs 267
    |||||::|||||||||::||||::||||::|||||||||||||::|||||||
```

```
301 GCAGCCTTCTGGGAGGTGTCGACTGCCAGTTGACCTGTGAGAACCAAAA 350
    |||||||||||||||:||||:||||||||||::|||||:|||:|||::||
267 nProCysPheProIleGlnLeuProGluAlaArgProAlaAlaGlyT 284

351 CCCCTGCGTTCCCCATCACAGCTTCCCTCAAACTCC......TCACGGACCA 394
    |||||||::||:||:|||||||||||||||||::: ::       |
284 hrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyThrGlyAsp 300

395 CTGCAGCTGCCTACCTTCTACCGCTCTCAGCCGCCTGGGCACTGGGGAC 444
    |||::|||||:||||||||||::||:||||::|||||||||||||||||
301 GlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArgGlnGlnMe 317

445 CAAGGAGCACTCTTTGGCAACCTGTCTCAGCCAATCCGAGGCAGCAGAT 494
    |||||||||||||||||||::|||:||:|||||||::||||||||||||
317 tAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGlySerSerP 334

495 GAATGGCGTTCCTCCTTGATGCTTCCACGGTGTATGGCAGCTCCC 544
    ||||||:||::|||||||||:||||||||||:::||||::|||||
334 roAlaLeuGluArgGlnLeuArgAsnTrpThrSerAlaGluGlyLeuLeu 350

545 CAGTGTTGGAGAAGCAGTTGCCAACTGGAGCAGCTCTCCAGGGCTGCTG 594
    |::::::||:||||:||:|||::||::|::||||::|:|||||||||||
351 ArgValHisGlyArgLeuAspAspGlyArgAlaTyrLeuProPheVa 367

595 CGTGTCAACACCCTCCACCTAGACTCTGGCCGTGCCTACCTGCCCTTC.. 642
    |||||||||||||||||||::|||::||:||||:|||||||::|:||
367 lProProArgAlaProAlaAlaCysAlaProGlnProGlyAsnProGlyG 384

643 ..........GCATTCAGCCGCCCTGTCTCCAGACGCCTGGTGCTCCACACG 682
             ||:||:||:|||::||||||||||||||||||:||||||||
384 luThrArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerGluVal 400

683 CCAACCGCACGCCCTGCTTCCTGGCTGGTGATGGCCGCGCCAGCGAGGTC 732
    ::||| |||||||||:||||||||:||||:||:||::|:||:||:||||
401 ProSerLeuThrAlaLeuHisThrLeuTrpLeuArgGluHisAsnArgLe 417

733 CCCGCCCTGCAGCAGTGCACACCTTGTGCGCTGCGCGCCGAGCACAACCGCCT 782
    ::|||     |||::|:::|||||||||::|||||:||::|:|||||||||
417 uAlaAlaAlaLeuLysAlaLeuAsnAlaHisTrpSerAlaAspAlaValT 434

783 GGCTACAGCTTTCAAAGCCATCAACACCATTGGAGCGCCAACACTGCCT 832
    |||::|::|:|||||||||||||||||||||||||:||::||||||:|||
434 yrGlnGluAlaArgLysValValGlyAlaLeuHisGlnIleIleThrLeu 450

833 ACCAGGAGGCACCCAAGGTGGTAGGGCATTGCACCAGATCATCACCATG 882
    |||||:||:||||:|||:||:|||::|||:|||:|||::||:|||:|||
451 ArgAspTyrIleProArgIleLeuGlyProGluAlaPheGlnGlnTyrVa 467

883 AGGGATTATATCCCCAAGATCTCGGGTCCCGATGCCTTCAGGCAGTATGT 932
    |:|||||||||:||:||||||:|||::|||::||||||:||::||||:||
467 lGlyProTyrGluGlyTyrAspSerThrAlaAsnProThrValSerAsnV 484

933 GGGCCCCATGAGGCTATAACCCCACGGTGAACCCTACTGTGTCCAACG 982
    |||::|::|:|::|:|||||||||||:||:||:|||::|:||:|||:||
484 alPheSerThrAlaAlaPheArgPheGlyHisAlaThrIleHisProLeu 500

983 TCTTCTCTACTGCTGCCTTCCGGCTTTGGCCATGGCCACATCCATCCACTG 1032
    ||||||:||::||||||||::|||::||||:||||||:||:|||:||||
501 ValArgArgLeuAsnSerPheGlnAspHisProAspLeuProGlyLe 517

1033 GTGAGACGACTAAACACGGACTTCCAGGACCACACAGAGTTCCACGGCT 1082
     ::||:|:|:|:||||||||:|||||:||:||::||||::|||::||:|
517 uTrpLeuHisGlnAlaPheSerProTrpThrLeuArgGlyGlyG 534

1083 GCAGCTGCGACGTGTCTTCTTCAGGCCCTGGAGGCTTATCCAGGAAGGTG 1132
     ||::||||:||:|:|||::||||||::|||||::||||:|||:|||||
534 lyLeuAspProLeuLeuArgGlyLeuLeuAlaArgProAlaLysLeuGln 550

1133 GTTTGGATCCAATAGTAAGGAGGCCTTCCTGGCAAGACCAAGCTGCAA 1182
     ||||||::||:|||::||:||||:|||::|||||||||::||::||||
550 lyLeuAspProLeuLeuArgGlyLeuLeuAlaArgProAlaLysLeuGln 550

1183 GTACAGGAGCAACTGATGAATGAGGAACTGACCGAGAGGCTCTTTGTGTT 1232
     ||||||||||::|||:|||||||||||::|||:|:|||:||:|||||||
551 ValGlnAspGlnLeuMetAsnGluGluLeuThrGluArgLeuPheValLe 567
```

DEFINITION   Halocynthia roretzi mRNA for homologue of mammalian thyroid
           peroxidase, complete cds.
ACCESSION   AB022197
VERSION   AB022197.1   GI:4587264
KEYWORDS   homologue of mammalian thyroid peroxidase.
SOURCE   Halocynthia roretzi endostyle tissue_lib:endostyle cDNA to mRNA,
           clone:HrTPO.
ORGANISM   Halocynthia roretzi
           Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
           Stolidobranchia; Pyuridae; Halocynthia.
REFERENCE   1 (sites)
  AUTHORS   Ogasawara,M., Di Lauro,R. and Satoh,N.
  TITLE   Ascidian homologue of the mammalian thyroid peroxidase genes are
           expressed in the thyroid equivalent region of endostyle
  JOURNAL   Unpublished (1999)
REFERENCE   2 (bases 1 to 3179)
  AUTHORS   Ogasawara,M., Di Lauro,R. and Satoh,N.
  TITLE   Direct Submission
  JOURNAL   Submitted (10-JAN-1999) to the DDBJ/EMBL/GenBank databases. Michio
           Ogasawara, Kyoto University, Department of Zoology, Graduate School
           of Science; Kitashirakawaoiwake-cho, Sakyo-ku, Kyoto, Kyoto
           606-8052, Japan (E-mail:ogasawara@ascidian.zool.kyoto-u.ac.jp,
           Tel:81-75-753-4095, Fax:81-75-705-1113)
FEATURES   Location/Qualifiers
     source   1..3179
           /organism="Halocynthia roretzi"
           /db_xref="taxon:7729"
           /clone="HrTPO"
           /tissue_lib="endostyle"
           /tissue_type="endostyle"
     CDS   66..2822
           /codon_start=1
           /product="homologue of mammalian thyroid peroxidase"
           /protein_id="BAA76689.1"
           /db_xref="GI:4587265"
           /translation="MSSNSLPSLSLGMLLVLFFVLSKIRNARCYEDHFQDTNRMQAEG
           DEFATKCMYEAMLNVNRAIRRTDIHWRKDTVWDSPMDLFQFFKRPPNEDAERMARASD
           IFETTVNYVNEKVLQKYRRMKRRVNATDVLSSKTLRDLAAFSGCIGNPLLKQCPDTCI
           ASKYRTTGQCNNLQNVYWGSSNHQLVRWQPSQYENGFSHPIGWNAETLRNNYRMPLV
           RKVSNDIIQTSNTNVTDDTDYSHMLVNGQYIDHFDFLTPQSLSTSTFQGLTNCQQTC
           RNEPPCCPILLPGEDSKRADAECLPFFRSSAVCGSGETSSLFNELKPEQMNAVTSFL
           DASTVYGSTDRMAYNLRNHTTDEGLMRVNDRFYDEGGRIFLPFNPNPCVQDQSDASG
           ERIPCFTAGDPRVSEHLTLSAIHTLWVRAHNIRARELKRINPHWYGETIYQEARKIVG
           SLHQIVHYKEYVPKIIGMTGMNLLGEYSEYNPSVNPTISNVFATAAFRFGHVTIAPIF
           RRLDGNFNEHPTHGNIFLHEAFFSPWRIIRQGLDPIFRGLIGRPAKLLTGTQIMHEE
           LREKLFQLQNKVALDLASLNLQRGRDHAIPLYSYWREFCNLTRVETFDELASEISDAS
           VELNWQNYTGHPGNLDLWLAGLVEDLVPGSRVGPTFLCLLTKQFQYLRDGDRFFYERV
           HTDEQIEELEKIRLANVLCYNSGLETVQRDVFSLAQYPDDFVRCSDLDPLNLEPWREE
           PEVGSCGRPQNIEYGDWQRCNDLVSYRCKMGYLDGQEELTCMDNGAFNAEPPMCVDV
           NECDQELKCNCEDICMNTVGSCRCMCSDGKILNEDGRTCSESPVTVVAPVGTNVAAIV
           TGVVLGVALLVLVVAVTYGVHKYTLLMQVATQAGTSSVNTVKMGISNSGFDSSSIDKT
           NMMH"
     polyA_site   3179
           /note="19 a nucleotides"
BASE COUNT   974 a   617 c   700 g   888 t
ORIGIN

alignment_scores:
     Quality: 1927.50   Length: 771
       Ratio: 3.323   Gaps: 10
  Percent Similarity: 75.227   Percent Identity: 48.119

alignment_block:
US-08-482-402A-3_COPY_1_848 x AB022197

Align seg 1/1   to: AB022197   from: 1   to: 3179

```
 76 PheSerLysLeuProGluProThrSerGlyValIleAlaArgAlaAlaGl  92
    ||::|:::       |||:::       ::|:||||||||||||||||||
321 TTCAAACAGACCCCCGAACGAAGATGCTGAGGATGGCGAGAGCTTCAGA  370

 92 uIleMetGluThrSerIleGlnAlaMetLysArgLysValAsnLeuLysT 109
```

567 uSerAsnSerSerThrLeuAspLeuAlaSerIleAsnLeuGlnArgGlyA 584
     ||::||| :::|||:||||||||:||
1233 GTCTAATGTGGCACCTTGGCATCCTGGCATCCACTGAACTTGCAGAGGGGCC 1282

584 rgAspHisGlyLeuProGlyTyrArgGluTrpArgGluPheCysGlyLeu 600
     |||||:||| |||||||||||||||:|||
1283 GGGACCACGGCTTACCAGGCTACAATGAGTGGAGAGAGTTCTGTGCCTTG 1332

601 ProArgLeuGluThrProAlaAspLeuSerThrAlaIleAlaSerArgSe 617
     |||||||| |||||:::|||||||
1333 TCACCCCTGGAGACACCAGCTGAGCTGACACAGGCCATTGCCAACAGAAG 1382

617 rValAlaAspLysIleLeuAspTyrLysHisProAspAsnIleAspV 634
     |||||:::||| :::|||||||
1383 CATGGTCAACAGATAATGAGTTATACAACGCATGCTGACAACATTGATG 1432

634 alTrpLeuGlyLeuAlaGluAsnPheLeuGluProArgAlaArgThrGly 650
     |||||||| ||||||||||||||
1433 TCTGGTTGGGAGGCCTGGCTGAAAGTTCTTACCCGGGAGCCCGCACCGGT 1482

651 ProLeuPheAlaCysTyrLeuIleGlyLeuLysGlnMetLysAlaLeuArgAspAspI 667
     ||:|||||| |||:|||||||
1483 CCCTCGTTTGCATGTATCATTGGACAGATGAAGGCTCTGAGGGATGG 1532

667 yAspTrpPheTrpIleGluAsnSerHisValPheThrAspAlaGlnArgA 684
     |||||:::|:|| |||||||||
1533 GGACAGGTTTGGTGGGAGAACAGCCATGTCTTCACAGACGCCTCAGAGGC 1582

684 rgGluLeuGluLysHisSerSerLeuSerArgValIleCysAspAsnThrGly 700
     :::|||||||| |||||||
1583 AGGAACTGGAAAAGCATTCACTACCTCGGGTCATCTGTGACAACACCGGC 1632

701 LeuThrArgValProMetAspAlaPheGlnValGlyLysPheProGluAs 717
     ||||||| |||||:::|||||||
1633 CTCACCAGATACCTGGATGCCTTCCGATTGGAAAGTTCCCCCAGGA 1682

717 pPheGluSerCysAspSerIleThrGlyMetAsnLeuGlnLeuAlaTrpArgG 734
     |||||||:::||| ||||||||
1683 CTTTGAATCCTGTGAGAAATCCCAGCATGGAACCTCAGACTGTGGAGGG 1732

734 luThrPheProGlnAspAspLysCysGlyPheProGluSerValGluAsn 750
     |||:::||| ||||||||||||||
1733 AGACCTTCCCACAGACGCACAAGTGTGTCTTCCCAGAGAAGGTGGACAAT 1782

751 GlyAspPheValHisCysGluGluSerGlyArgArgValLeuValTyrSe 767
     ||:::|||::: |||:||||||
1783 GGGAACTTTGTGCACTGTGAAGAATCTGGAAGCTGGTACTGGTGTATTC 1832

767 rCysArgHisGlyTyrGluLeuGlnGlyArgGluGlnLeuThrCysThrG 784
     ||:|| |||||||:::|||||||
1833 CTGTTCCATGATCAAGCTGCAAGGCCAGGAGGCAGGTCACATGTACCC 1882

784 lnGluGlyTrpAspPhePheGlnProProLeuCysLysLysAspValAsnGluCys 800
     ||||||| :::|||||||
1883 AGATGGGATGGGACTTCAGAGCCTCAGAGCTGTGTAAAGATGTTAATGAGTGT 1932

801 AlaAspGlyAlaHisProProCysHisAlaSerAlaArgCysAsnTh 817
     ||:::||| ||||||||||||
1933 GCAGATCTGACACACCCACCTTGCCACTCCTCCGCAAAGTGCAAGAACAC 1982

817 rLysGlyGlyPheGlnCysLeuCysAlaAspProThrGluLeuGlyAspA 834
     ||:::||| |||||||||||
1983 CAAGGGAGGTTCCAGTGTGTGTGCACAGACCCCTACATGCTAGGTGAGG 2032

834 spGlyArgThrCysValAspSerGlyArgGluArgValThr 848
     |||||||||||
2033 ATGAGAAGACCTGCATAGATTCTGGCAGGTCACCTCGGGCATCC 2076

Mon Nov 26 15:43:03 2001

us-08-482-402a-3_copy_1_848.rge

```
1253 TTGTTTACTGCGGGAGACCCAAGAGTTTCAGAACACATTGACTCTTATCAG 1302
 405 laLeuHisThrLeuTrpLeuArgGluHisAsnArgLeuAlaAlaAlaLeu  421
     |:::||||||||| |||||||  |||||| ||||::||||
1303 CAATTCACACTTTATGGGTAAGAGCACACAACAGAATTGCTCGTGAATTG 1352
 422 LysAlaLeuAsnAlaHisTrpSerAlaAspAlaValTyrGlnGluAlaAr 438
     |||  :::|||  ||||||::|||  ||| |||||::||||||
1353 AAAAGAAATCAACCCTACTGGTACGGGAAACGATTTATCAAGAGGCAAG 1402
 438 gLysValValGlyAlaLeuHisGlnIleIleThrLeuArgAspTyrIleP 455
     ||:|||||| ||||||||||:::|||||  ||||||||||:::|||
1403 GGAGATCGTGGGCAGTCTTCATCAAATGTCCAATTACAAGAAATATGTTC 1452
 455 roArgIleLeuGlyProGluAlaPheGlnTyrValGlyProTyrGlu    471
     |:::|||||  ::|||  ::::|||
1453 CAAAGATTATCGGAATGACTGGA...ATGAACTGCTCGGAGAATATTCC 1499
 472 GlyTyrAspSerThrAlaAsnProThrValSerAsnValPheSerThrAl 488
     |||::::||||||||||||||:::||||
1500 GAATACACCCGAGCGTCAACCGGTCCACCATTTCAAACGTTTTCGCCACGGC 1549
 488 aAlaPheArgPheGlyHisAlaThrIleHisProLeuValArgArgLeuA 505
     |||||::||||||||||||||::::|||
1550 CGCATTCCGATTCCGTCATGTGACAATAGCACACCGGAAATTCCGGAGGTTGG 1599
 505 spAlaSerPheGlnGluHisProHisProGlyLeuTrpLeuHisGln    521
     |||:::|||||||||::::|||
1600 ATGGAAATTCAACGAACATCCGACCACACCGAAACATCTTTTGCATGAG 1649
 521 AlaPhePheSerProTrpProThrLeuLeuArgGlyGlyLeuGlyAspPro 538
     |||||||||||||||||::::|||
1650 GCATTCTTTTCGCCCTGGACAATCATACGGCAGGTGGACTCGATCCGAT 1699
 538 uIleArgGlyLeuLeuAlaLeuArgProAlaLysLeuGlnIleGlnAspGlnL 555
     |||||||||||||||::::|||||||||||::::|||
1700 CTTCCGGGGATTGATTGGAGAACCTGCCAAACTAATAACTGGTACACAAA 1749
 555 euMetAsnGluGluLeuThrGluArgLeuPheValLeuSerAsnSerSer 571
     ::|||::::|||||||||||||||
1750 TCATGCACGAAGAACTGCCAGAAAAGTTGTTCCAGTTGCAAAATAAGGTT 1799
 572 ThrLeuAspLeuGlnAlaSerIleAsnLeuGlnAsnLeuGlyArgAspHisGlyLe 588
     :|||||||||||||||||||||||||||||:::|||
1800 GCCCTGATCTTGCTTCACTGAACCTCCACGAGGTCGTGACCACGCTAT 1849
 588 uProGlyTyrAsnGluTrpArgGlnTyrLeuProArgLeuGluT      605
     |:::|||||||||||::::|||
1850 CCCACTTTATAGTACTGGAGAGATTTTGTAACTGACACAGAGTGAAA 1899
 605 hrProAlaAspLeuSerThrAlaIleAlaSerArgSerValAlaAspLys 621
     ||  :::||||||||||::::|||
1900 CATTTGACGAATTGGCATCGGAAATAAGGCGATGCCTCGGTAGAGCTAAAC 1949
 622 IleLeuAspLeuTyrLysHisProAspAsnIleAspValTrpLeuGlyLyLe 638
     :|||||||||||::::||||::::|||
1950 TGGCAGAATTATACGGGTCATCCTGGTAATCTGGATTTGTGGCTTGCTGG 1999
 638 yLeuAlaGluAsnPheLeuProArgAlaArgThrGlyProLeuPheAlaC 655
     |::::||||||||||||::::|||
2000 CTTAGTTGAGGATCTTGTCCCTGGAAGTCAGTCGGACCACCAACTTTCCTAT 2049
 655 ysLeuIleGlyLysGlnMetLysArgArgAspAspGlyAspTrpTrpTrp 671
     ||  :::|||||||||||||::::|||
2050 GCCTTCTTACCAAGCAGTTCCAGTATCTCAGAGATGGTGATAGGTTTTC 2099
 672 TrpGluAsnSerHisValPheThrAspAlaGlnArgArgGluGluLeuGluLy 688
     :::|||  |||||||::::||||
2100 TATGAGCCGTTCAC......ACCGATGAACAGATTGAGGAATTGGAGAA 2143
 688 sHisSerLeuSerArgValIleCysAspAsnThrGlyLeuThrArgValP 705
     |||  ::::||||||:::||::::
2144 AATTCGTCTCGCCAACGTTTTGTGTTACAATAGTGACTTGAAACAGTTC 2193
```

```
 371                     :|||||:::||||||:::::::|||||  ||| 420
     TATATTCGAAACAACGGTCAATTACGTTAATGAGAAAGTTCTTCAGAAAT
 109 hrGlnLnSerGln........HisProThrAspAlaLeuSerGluAsp 122
     ::::|| :::
 421 ACAGAGAAATGCAAACGTCGTAAATGCAACGATGTTCTTCTTCAAAA 470
 123 LeuLeuSerIleIleAlaAsnMetSerGlyCysLeuProTyrMetLeuPr 139
     :::::|||  ||||||||||||||||||||
 471 ACATTGAGAGATCTTGCTGCGTTTCTCGATGCATTGGAAATCCTCTATT 520
 139 oProLysCysProAsnThrCysLeuAlaAsnLysTyrArgProIleThrG 156
     ::||||||||||::::||||||:::||||||  |||||||
 521 GAAACAATGTCCAGATACTTGTATTGCATCAAAATACAGGACCATAACTG 570
 156 lyAlaCysAsnAsnArgAspHisProArgTrpGlyAlaSerAsnThrAla 172
     |:|||||||  ||||||:::||||
 571 GCCAATGCAATAACTTACAAATGTACCTGGGGATCATCAAATCATCAA 620
 173 LeuAlaArgTrpLeuProProValTyrGluAspGlyPheSerGlnProAr 189
     :::||||||||||||  |||||||||
 621 CTCGTACGGTTGGCAGCCGCCGTCTCAATATGAATGCAGTCAGTCATCCCAAT 670
 189 gGlyTrpAsnProGlyPheLeuTyrAsnGlyPheProLeuProProValA 206
     |||  :::|||  ||  :::
 671 CGGATGGAGTCGGAAAACACTTAGAAACAACTATCGTATGCCTTTGGTTC 720
 206 rgGluValThrArgHisValIleGlnValSerAsnGluValValThrAsp 222
     |||| |||||||  ||||||||
 721 GAAAAGTTTCCAACCAATATCATTCAGACACAACTACTAATGTTACAGAC 770
 223 AspAspArgTyrSerAspLeuMetAlaTrpGlyGlnTyrIleLeuAspPhi 239
     |||  :::||||||||||
 771 GATACTGATTACTCTCATATGCTTGTGTGTGGGGCCAGTATATCCACCA 820
 239 sAspIleAlaPheThrProGlnSerThrSerLysLysAlaAlaPheGlyLyG 256
     |||::::  :::|||
 821 TGATTTTGACTCCACAAGGTTGAGTACGGTCTACCTTTCAAGGAC 870
 256 lySerAspCysGlnMetThrCysAsnProCysPheProIle          272
     :::||||||:::::::::||| ||||||||||||||
 871 TGACAAATTGCCAACAACTTGCCACAACTTGCCTTGCTTTCCTATA 920
 273 GlnLeuProGluGluArgProAlaAlaGlyThrAlaCysLeuProPh 289
     :::   ||| |||||
 921 CTGCCTTCCAGGCGGATTCGAAAAGGGCCGATGCCGAATGTCTCCCATT 970
 289 eTyrArgSerSerAlaAlaCysGlyThrGlyAspGlnGlyAlaLeuPheG 306
     ::::|||||||||||  :::::||||
 971 CTTCAGATCTTCAGCTGTGACTGCGTGAGTACGGTGGAGAAACATCCAGTCTATTTTA 1020
 306 lyAsnLeuSerThrAlaAsnProGlnGlnMetAsnGlyLeuThrSer 322
     :::::||||||||||::::||||  :::
1021 TGACAAATTGCCACAACTGCCAGAAATGAACAGAACAAATGCTGTGACTTCA 1061
 323 GlnLeuProAlaSerProAlaSerThrValTyrGlySerProAlaArgGl 339
     ::
1062 TTCCCTTGACGGCAGTGTATCGGAAGCACGGAATGGCTACAA 1111
 339 nLeuArgAsnTrpProSerAlaGluGlyLyLeuLeuArgValHisGlyArgL 356
     ||||||:::||||||:::||| :::
1112 TCTTAGAAATCATACTACAGATGAGGCTTGATGCGCGTAAATGATAGAT 1161
 356 ySerGlyArgAlaTyrLeuProPheValProProArgAla          371
     ::::|||||  :::|||  |:::
1162 ACGAGCTC........AAACCAAGAACAAACAAATGAATGCTGTGACTTCA 1208
 372 ProAlaAlaCysGlyAspProGluProGlyAlaProGlyThrArgGlyPr 388
     :::   ||| :::
1209 ......CCAATGTGTACAAGATCAGTCCGATGCTAGCGCGGAAAGAATACC 1252
 388 oCysPheLeuAlaGlyLeuLeuAlaGlySerGluValProSerLeuThrA 405
     :::|||||||:::||:::
```

/product="homologue of mammalian thyroid peroxidase"
/protein_id="BAA76688.1"
/db_xref="GI:4587263"
/translation="MWTKCIAVVLFCKLVFADEQVFPDETGYEGDEFAREALREAFQQ
VRRAIFRTDKSWSDEKVWRNPMTLFQYFKRPPSESSDMAKAGDIYHCAFDIVHDKVS
NRLSRFKRRVNVTDVLSSATIAELAKFSGCRANGVSPTCPDTCIDNMYRTHTGCNNK
VNRYWGASNNPVRWRPAQYENEFSTPIGWDSQRSYNGYQLPVVRKVSNDIMRTSNTR
VTGDPLYSHMLVVWGQYIDHDLDFTPQSLSTSTFQGLZTDCKKTCKNESPCYPMQVPSD
DPRITTASCLPFFRSAAVCGTGDTSSLFHSIRPREQINAVTSFVDASTVYGSTDSLNR
ILRNLTNDDGLMKVNTMFKDGNWDYLPFDPNNPCVQDPFDASGVNIPCFHAGDGRVSE
HITLSAIHTVWVREHNRIARMLKSMNPHWSGEIIYQEARKIVGAYHQIVHWKEYVPKI
IGPAGLRMGNVTGYRENENPTVSNVFATAAFRFGHATISPQFRRLDENFNNHPQFPT
ILLHEAFFSPWRMIREGGMDPILRGLIGRPAKLIKADEMMHEELRDKLFALQNQVALD
LASLNLQRGRDHALPLYNDWRECGLARANNFSDLAGEIKDKAIRDKLEALYGHPGNI
DLWLAGLSEDLMDGSRGPVFTCLLARQFKFLRNGDRFYYENPNVFTPNQVTALNRLS
FARVLCDNSGLTRVQPDLFMLRDTSQFVDCANIPNLDLSQWREDPAVGTCGTPPSISN
GWWKKCGGVSYRCMSGYQLAGPNEVQCMNSAFTAIPSCVDINECDAQNGGCSDRCMN
TAGSYICVCNDDRMLGSDGKTCMATTVQVTPSPAVMGVTATTPANVTAIVVGSILGAA
IFLAFCVICYLVYKNTKLHTKSKYHNTGKFSLSDPSAYDNPSAVMDNDSQTKM"

polyA_site   2915
/note="20 a nucleotides"

BASE COUNT   833 a   610 c   687 g   785 t
ORIGIN

alignment_scores:
    Quality: 1879.00   Length: 775
      Ratio: 3.302   Gaps: 11
Percent Similarity: 73.419   Percent Identity: 46.323

alignment_block:
US-08-482-402A-3_COPY_1_848 x AB022196

Align seg 1/1   to: AB022196   from: 1   to: 2915

```
 76 PheSerLysProGluProThrSerGlyGlyValIleAlaArgAlaAlaGl  92
    |||::::||||::::||:::::||||                  MetGluThrSerI  98
232 TTTAAACGACCGCCCAGCGAAGAATCGAGCGATATGGCAAAGCTGGGGA 281

 92 uIle..........................................  98
    :|||
282 TATTTATCATTGTGCGTTTGATATTGTACATGACAAGGTGTCGAACCGAC 331

 98 leGlnAlaMetLysArgLysValAsnLeuLysThrGlnGlnSerGlnHis 114
    :||||::|||::||||::||||
332 TTTCAAGGTTAAACGCCGGGTAAACGTC................. 360

115 ProThrAspAlaLeuSerGluAspLeuLeuSerIleIleAlaAsnMetSe 131
    :||||||:||||||:|||          |||||
361 ..ACTGATGTGCTGTCAAGCGCAACGATCCAGAGTTGGCCAAATTTTC 407

131 rGlyCysLeuProTyrMetLeuProProLysCysProAsnThrCysLeuA 148
    ||||||   :::  ||| ||||||::||||::|||||
408 TGGTTGTCGTGCAAATGGTGAGTCGAACATGTCCGGATACTTGTATCG 457

148 laAsnLysTyrArgProIleThrGlyAlaCysAsnAsnArgAspHisPro 164
    |||||  ||||||  ||||||:|||::    :|||
458 ACAATATGTACAGAACCCACCACCGTGAATGCAACAACAAGTTAATCGA 507

165 ArgTrpGlyAlaSerAsnThrAlaLeuAlaArgTrpLeuProProValTy 181
    |||||||||  |||  ||||:||||  |||  ||||::
508 TACTGGGGTGCTTCGAATAATCCGTTCGTTAGATGGGCGCCCTGCACAATA 557

181 rGluAspGlyPheSerGlnProArgGlyTrpAsnProGlyPheLeuTyrA 198
    ||||:::  :|||||  |||  ||||||
558 CGAAAATGAATTCAGCACTCCAATAGGTTGGGATAGCCAACGTTCGTATA 607

198 snGlyPheProLeuProProValArgGluValThrArgHisValIleGln 214
    :|||||:  ||||||:  |||||:::||| |||||:||||
608 ATGGATACCAACCAATTACCTGTTGTCGCAAAGTTCAAACGATATAATGAGA 657

215 ValSerAsnGluValValAlaThrAspAspAspAspArgTyrSerAspLeuLeuMe 231
    ::|||||| ||| |||||| ::|||::
```

```
705 roMetAspAlaPheGlnValGlyLysPheProGluProAspPheGluSerCys 721
    |||:::|||:::::||||::||| |||:::|||
2194 AACGTGACGTATTCCACTTGCGCAGTATCCTGATGATTTTGTCCGATGC 2243

722 AspSerIleThrGlyMetAsnLeuGluAlaTrpArgGluThrPhePro Gl 738
    :::::::  |||||||||  |||||||
2244 TCTGATTTGGACCCGTTGAACTTGGAGCCATGGAGAGAAGAA...CCAGA 2290

738 nAspAspLysCysGlyPheProGluSerValGluAsnGlyIleAspPheValH 755
    :  ::|||||  |||::::||| ||||:::  :
2291 AGTTGGTTCATGTGGAAGGCGCAAACATTGAGTATGCGACTGCAAA 2340

755 isCysGluSerGlyArgGlyArgValLeuValTyrSerCysArgHisGly 771
    :::||||||::|||  ::::: ||| |||:::: |||
2341 GGTGCAATGAC.........CTCGTCAGTTATAGATGCAAAATGGGT 2378

772 TyrGluLeuGlnGlyArgGluGlnLeuThrCysThrGlnGluGly...Tr 787
    ::: ||||::::|||||||::|  :::::|| |||||||
2379 TTTTACTTGGATGATGATGGACCAAGAGGAGTTAACATGTATGGATAATGTGCTTT 2428

787 pAspPheGlnProLeuCysLysAspValAsnGluCysValAlaAspGlyA 804
    ::: :::|||||||::|||:::||| ||||||||||||||
2429 TAATGCAGAACCACCAAATGTGTTGATGTGAACGAATGT.......GACC 2472

804 laHisProCysHisAlaSerAlaArgCysArgAsnThrLysGlyGly 820
    ||:::::: |||||||||| |||||| ||||||
2473 AGGAATTGAAATGTTGAGGATATATGCATGACACTGTTCGTAGC 2522

821 PheGlnCysLeuCysAlaAspProTyrGluLeuGlyAspAspGlyArgTh 837
    :::::||||||::::|||| |||||||||||||||
2523 TGTCGATGATGTGTTCTGATGGCAAAATATTAAACGAGGATGAAGAAC 2572

837 rCysValAspSer 841
    |||| :::|||
2573 TTGTTCTGAGTCC 2585
```

seq_name: gb_in:AB022196

us-08-482-402a-3_copy_1_848.rge

```
658  ACTTCGAATACAAGAGTTACAGGCGATCCTTCGTATTCCCATATGCTGGT  707
231  tAlaTrpGlyGlnTyrIleAspHisAspIleAlaPheThrProGlnSerT  248
     :::|||||||||||||||||||||||::: |||||||||||||
708  AGTGTGGGACAATACATTGACCATGATTTAGATTCACTCCACAAAGTT     757

248  hrSerLysAlaAlaPheGlyGlyGlySerAspCysGlnMetThrCysGlu  264
     ||||||||||||||||||||||||   ||||||
758  TAAGTACCTCCACTTTCAAGGATTAACTGATTGCAAGAGACGTGTAAA    807

265  AsnGlnAsnProCysPheProIleGlnLeuProGluGluAlaArgProAl  281
     |||||||::::|||||::||||||||||||||
808  AACGAGTCACCGTGTATCCAATGCAGGTACCTTCTGACGACCCTCCAT    857

281  aAlaGlyThrAlaCysLeuProPheTyrArgSerSerAlaAlaCysGlyT  298
     :||||||||||||||||||||::::::|||::|||||
858  CACAACAGCATCTTGCTCTCCATTTTCCGTTCACGAGCTGTGTGTGGGA   907

298  hrGlyAspGlnGlyAlaLeuPheGlyAsnLeuSerThrAlaAsnProArg  314
     ||||||||     ||||     :::::::||
908  CAGGTGACACAGTCACTCACTATTTCATTCCATC........AGACCACGT 948

315  GlnGlnMetAsnGlyLeuThrSerPheLeuAspAlaSerThrValTyrGl  331
     ||||||:::|||||||||::||||||||||:::
949  GAGCAAATCAACGCTGACGTCATTTGTCGACGCAAGCACGGTATACGG    998

331  ySerSerProAlaLeuGluArgGlnLeuArgArgAsnTrpThrSerAlaGluG 348
     |||||:::||:::|||   |||||||||   :::
999  AAGTACTGATTCACTGAATCGAATTTACGTAATTTGACCAACGATGATG   1048

348  lyLeuLeuArgValHisGlyArgLeuArgAspSerGlyArgGlyAlaTyrLeu 364
     |||||||:::|||       ::||||       ||||||
1049 GTCTGATGAAAGTAAATACCATGTTAAAGACGGCAATTGGGATTATTTG   1098

365  ProPheValProProAlaProAlaAlaCysAlaProGluProGlyAs    381
     ||||||       ||       ::
1099 CCATTCGATTCCGAACAAC..........CCGTGTGTACAAGATCCTTTCGA 1139

381  nProGlyGluThrArgGlyProCysPheLeuAlaGlyLysAspGlyArgAlaS 398
     ::       ||||||||||||||:::|
1140 TGCGAGGCCGGCGTCAACATTCCTTGTTTCCACGCAGGAGATGGAAGAGTAA 1189

398  erGluValProSerLeuThrHisThrAlaLeuHisThrLeuTrpLeuArgGluHis 414
     |||||       ::|||||||||||:::|||||||||
1190 GCGAGCATTTAACATTAAGTGCAATACACACCAGTATGGGTGCGAGACAC   1239

415  AsnArgLeuAlaAlaLeuLysAlaLeuLysAsnAlaHisTrpSerAlaAs  431
     |||||||||||||||       ::|||
1240 AACAGGATTGCTCGTATGTTAAAGTCCATGAACCCACTGTCCGGGTGA    1289

431  pAlaValTyrGlnGluAlaAsnArgLysValValGlyValAlaLeuHisGlnIleI 448
     :::||||||||||||||||||||:::|
1290 AATTATTTACCAGGAAGCAAGGAAGATAGTTGGAGCTTATCACCAGATCG   1339

448  leThrLeuArgAspTyrIleProArgIleLeuGlyLeuGlyProGluAlaPheGln 464
     ::||||||||||||||||||||||
1340 TGCATTGGAAGGAATACGTGCCAAAGATAATCGGCCCAGCTGGTTGCGC   1389

465  GlnTyrValGlyTyrProTyrGluGlyTyrTyrAspSerThrAlaAsnProThrVa 481
     |||   ::|||   ||||||
1390 ATG...ATGGGCAACTATACAGGTTATAGAGAAGAATGAAAAACCCACAGT  1436

481  lSerAsnValPheSerThrArgAlaAlaPheArgPheGlyHisAlaThrIleH  498
     :::||||||||||||||||||||||||||||||
1437 TTCAAACGTTTCGCCACAGCTGCGTTTCGTTTCGGCCATGCCGACTATCT   1486

498  isProLeuValArgArgLeuAspLeuSerPheGlnGluHisProAspLeu    514
     ||||||       :::::||||||
1487 CGCCCCAGTTCAGGAGACTGGACGAAAACTTAACAATCACCCGCAGTTC   1536

515  ProGlyLeuTrpLeuHisGlnAlaPheSerProTrpProThrLeuLeuAr  531
     |||||||       ::::||       ||||
1537 CCGACCATTCTTTTGCATGAAGCTTTTTTCTCCCCATGGGCAATGATAAG  1586
```

```
531  gGlyGlyLeuAspProLeuIleArgGlyLeuLeuAlaAlaArgProAlaL  548
     |||||||||||||||||||||:::||||||       ::||||||
1587 GGAGGGGGAATGGATCCTATATTGAGAGGTTTGATCGGCCGACCGGCCA   1636

548  ysLeuGlnValGlnAsnProLeuMetAsnGluGluLeuThrGluArgLeu  564
     |||||       ||   |||||||       :::::||||||
1637 AACTTATCAAAGCTGACGAAATGATGCACGAAGAGCTTAGGGATAAACTC  1686

565  PheValLeuSerAsnSerThrLeuAspLeuAlaSerIleAsnLeuGl    581
     :::||::::|||       ::||||||||||||||||
1687 TTTGCTTTACAAAACCAGGTGGCCATTGGATCTTGCCTCACTGAACTACA  1736

581  nArgGlyArgAspHisGlyLeuProGlyTyrArgAsnGluTrpArgGluPheC 598
     ||       ||||||||||||||||||||||
1737 GAGAGGTCGTGACCACGCGTTACCATTATACAATGATTGGCGTGAGGGAGT 1786

598  ysGlyLeuProArgLeuGluLeuThrProAlaAspLeuSerThrAlaIleAlaAla 614
     ||||||       ::||||       |||
1787 GTGGTTTAGCAAGGCGGACAACTTCAGTGACTTGGCTGGTGAAATAAAA   1836

615  SerArgSerValAlaAlaLysIleLeuAspLeuTyrLysHisProAspAspAs 631
     :::||||||||       ||||||       |||||
1837 GACAAAGCAATTCGTGATAAACTAGAAGCTCTATACGGTCATCCAGGAAA  1886

631  nIleAspValTrpLeuGlyGlyLeuAlaGluAsnPheLeuProArgArgAlaA 648
     ||:::||||||||||||||||||||||||||
1887 CATAGATTTATGGTTAGCTGGTCTTTCCGGAGATTTAATGCGACGGTTCGC 1936

648  rgThrGlyProLeuPheAlaCysLeuIleGlyLysGlnMetLysAlaLeu   664
     ||       :::||||||||||||||||       :::||
1937 GTGGTGGACCAGTCTTCACTTGCCTGCTTGCTCAGTTTAAATTCCTT     1986

665  ArgAspGlyAspTyrProPheTrpProGluAsnSerHisValPheThrAspAl 681
     ||       ::||       :::|||||||||
1987 CGAAACGGGAGACAGATTTTACTACGAGAATCCAAATGTTTCACACCGAA  2036

681  aGlnArgArgGluLeuGluLysHisSerLeuSerArgValIleCysAspSerA  698
     ::||||||       :::||||       ||||
2037 TCAGGTCACAGCGTTAAACAGATTATCCTTCGCCCCGTGTGTTGTGATA   2086

698  snThrGlyLeuThrArgValProMetAspAlaPheGlnValGlyLysLysPhe  714
     ||:::||||||||||||||       |||
2087 ACTCTGGTTTAACGCAGTCCAACCCGACTTGTTCATGCTA...CGTGAT   2133

715  ProGluAspPheGluSerCysAspSerIleThrGlyMetAsnLeuGluAl   731
     ::::|||       ::|||       ::||||||
2134 ACATCTCAGTTTGTTGATTGTCAAACATTCCTAACCTGGATCTTTCACA   2183

731  aTrpArgGluThrPheProGlnAspAspLysCysGlyPheProGluSerV   748
     |||       ||||       ::::||||||||||||||
2184 GTGGCGTGAAGAC...CCAGCAGTTGGTACATGCCGGTACACCACCTAGTA 2230

748  alGluAsnGlyAspPheValHisCysGluSerGlyGluSerGlyArgArgValLeu 764
     :::||||||||||||||||||||||||||
2231 TAAGTAACGGTTGGTGGAAGAAGTGTGGTGGGGGT.........GTT     2268

765  ValTyrSerCysArgHisGlyTyrGluLeuGlnGlyLeuArgGluGlnLeuThr  781
     ||||||       :::|||       ::||||||
2269 AGTTATCGTTGTATGTCTGGATACCAACTTGCTGACCAAATGAGGTGCA   2318

781  rCysThrGlnGluGlyTyrProLeuProLeuCysLysAspAspValA      798
     |||       :::||||||       ::::|
2319 ATGTATGAACTCAGCTTTTACC...GCCATACCATCTTGTTGACATAA    2365

798  snGluCysAlaAspGlyAlaHisProProCysHisSerAlaSerAlaArgCys  814
     ||||||       :::       ::||||||||||||||
2366 ACGAGTGC...GATGCACGAACGGTGGGTGC.......AGTGACAGGTGT   2406

815  ArgAsnThrLysGlyGlyPheGlnCysLeuCysAlaAspProTyrGluGlu   831
     ||||||       :::||       |||
2407 ATGAACAACCGCTGTGGTAGTTACATTTGTGTTGCAATGACGATCGTATGCT 2456
```

seq_name: gb_ov:AF349034

```
seq_documentation_block:
LOCUS           AF349034        3846 bp    mRNA          VRT       02-AUG-2001
DEFINITION      Danio rerio myeloperoxidase mRNA, complete cds.
ACCESSION       AF349034
VERSION         AF349034.1  GI:15077546
KEYWORDS        .
SOURCE          zebrafish.
ORGANISM        Danio rerio
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
                Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE       1  (bases 1 to 3846)
AUTHORS         Langenau,D.M., Bennett,C.M., Deng,M., Kanki,J.P., Rhodes,J.,
                Liu,T.X., Delahaye-Brown,A. and Look,A.T.
TITLE           Genetic Analysis of Myelopoiesis in the zebrafish
JOURNAL         Unpublished
REFERENCE       2  (bases 1 to 3846)
AUTHORS         Langenau,D.M., Bennett,C.M., Deng,M., Kanki,J.P., Rhodes,J.,
                Liu,T.X., Delahaye-Brown,A. and Look,A.T.
TITLE           Direct Submission
JOURNAL         Submitted (13-FEB-2001) Pediatric Oncology, Dana-Farber Cancer
                Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES        Location/Qualifiers
     source     1..3846
                /organism="Danio rerio"
                /db_xref="taxon:7955"
     CDS        73..2361
                /codon_start=1
                /product="myeloperoxidase"
                /protein_id="AAK83239.1"
                /db_xref="GI:15077547"
                /translation="MNLPTFLFVVGCCALSVGAEEESGRPPFILDSIEEAKKIVDAA
                YKYSRDESLARVRKDVIKPSDKLRLLKQPARKTREAVRAADYMAQTLRLISERAHHVH
                KRSINATNLLTQDELQIITRMTCEAQTRPPSCRTTPLINKYRTASGVCNNRKNPLLG
                ASNTPFARWLPAQYEDAVSQPKGWDPNKLYNGAALPMVRLVSNRILATADADIESDHD
                FTPMLITIFGQWVDHDLIFTPFSPSIRSFSNGIDCENSCERSEPCFPISAPPGQORLRP
                NTCLPVFRSAPTCGSGHAYMFGEVPNVREOINTLTAYLDAGQYYGSEDGLAKELRDL
                TNDGGLLRVNNRFKDNGRELLPFTSVNTNLCATRQKILNDSTLTEVPCFIAGDARVDE
                NPALNSLHTLFVREHNRLARALHLVNLNPTWSSETLYQEARKIVGAFNQILVIKEYLPLI
                VGPDAYNRHLGPYGNENVDPTIANVFATAAFRFAHLTIQPFIFRLDENYKNHPQFP
                SVPLYEAFFSPWRVIFEGGIDPVLRGLIGRPAKLNTQDHMLVNALRERLFAFTSHIAL
                DLASLNMQRGRDHAIPGYNAWRRFCGLSAPONEQELAVVMNNTKLARKLIELYGTPEN
                IDVWLGGVAEPFAPGGRVGSLFACLISRQFOKIRDGDRLWFESNGVFTTKQKTALASV
                SMARIICDNTGLLKVPRDPFRFRSPASFVNCGNIPAFDLEPWKETGDGGINISDIGNE
                IEIPSQDLPDPGDLQENVVHRV"
BASE COUNT      1062 a      850 c      826 g     1108 t
ORIGIN
```

```
alignment_scores:
       Quality:  1670.50          Length:    855
         Ratio:     2.808           Gaps:     15
Percent Similarity:  69.591    Percent Identity:  40.234

alignment_block:
US-08-482-402A-3_COPY_1_848 x AF349034

Align seg 1/1  to: AF349034  from: 1  to: 3846
```

```
 32 GlyLysProGluSerArg..........ValSerSerValLeuGl 44
    |||||||||                     ::  |||   :::||
127 GGGCGAGAAGAAGAAAGTCCTGGAAGACCTTTATTCTTGATTCAATTGA 176

 44 uGluSerLysArgLeuValAspThrAlaMetTyrAlaMetThrMetGlnArgA 61
    ||||::|||::|||||||                :::  ::::||
177 GGAAGCAAAGAAAATAGTGACACGCTGCTTACAAGTATTCTCGAGATGAGA 226
```

```
 61 snLeuLys.....LysArgGlyIleLeuSerGlyAlaGlnLeuLeuSer 75
    :::||             ::::: ::::|| ::::|   |||
227 GTTTAGCAAGGGTCGCAAAGATGTCATTAAGCCTTCTGACAAGCTACGT 276

 76 PheSerLysLeuProGluProThrSerGlyValIleAlaArgAlaAlaGl 92
    ::: ||||                          :::  |||:||:||
277 CTGCTGAAACAGCCAGCTCGAAAGACACGGAGGCCGTAGAGAGCTGCAGA 326

 92 uIleMetGluThrSerIleGlnAlaMetLysArgLysValAsnLeuLysT 109
    :  |||||                ||| ||||    ::::|||||||
327 CTACATGGCACAAACGCTGAGACTGATCAGTGAGAAAGCCCATCATGTAC 376

109 hrGlnGlnSerGlnHisProThrAspAlaLeuSerGluAspLeuLeuSer 125
    ::::|||   |||          ||| :::|||  |||:||   :::||
377 ACAAGAGGTCCATCAATGCCACAAACCTGCTCACTCACTGACGAGCTTCAA 426

126 IleIleAlaAsnMetSerGlyCysLeuProTyrMetLeuProProLysCy 142
    ||||||                 |||||||||::::::| :::||||
427 ACCATTACAGTATGACAGGTCTGTGAAGCTCAAACCGGCCACCATCTTG 476

142 sProAsnThrCysLeuAlaAsnLysTyrArgProIleThrGlyAlaCysA 159
    | ||||||||||           |||||||||||  :::  :::|||
477 TAGGACCACCACCCCTCATAAACAGTATCGAACTGCCAGCGGTGTCTGCA 526

159 snAsnArgAspHisProArgTrpGlyAlaSerAsnThrAlaLeuAlaArg 175
    ||||||||    :::|||         ||||||||||||||  ::|||||
527 ACAATCGGAGAACCCCACTTCTTGGTGCATCCAACACTCCATTGCCCGC 576

176 TrpLeuProProValTyrGluAspGlyPheSerGlnProArgGlyTrpAs 192
    ||||||||                |||||| ::::|||||||||:::||||
577 TGGTTGCCCGCTCCATATGAGGACGCCGGTTTCTCAGCCAAAGGATGGGA 626

192 nProGlyPheLeuTyrAsnGlyPheProLeuProProValArgGluValT 209
    :::|||                  |||:::|||||||||||:|||||
627 TCCCACAAGTTGTACAATGGTGCAGCGCTGCCCATGGTTAGACTGGTTT 676

209 hrArgHisValIleGlnValSerAsnGluValValThrAspAspArg 225
    :::::::||||||||               :::     |||::||||
677 CCAATCGCATTCTAGCCACTGCAGATGCTGACATAGAGAGTGACCATGAC 726

226 TyrSerAspLeuGluMetAlaTrpGlyGlnTyrIleAspHisAspIleAl 242
    :::|||||    ::: ||| |||            ||||||:::||||||
727 TTTACCTTCATGCTCACCACCATTTTGGACAATGGGTGACCACGATCGAC 776

242 aPheThrProGlnSerThrSerLysAlaAlaPheGlyGlyGlySerAspC 259
    ||| |||         |||    :::||| ||||||||||||||||
777 TTTTACACCTTTCTCACCAAGTATTAGGTTCCTTTAGTAATGGCATTGACT 826

259 ysGlnMetThrCysGluAsnGlnAsnProCysPheProIleGlnLeuPro 275
    ::: |||                 ::::::||||||||||||||||||
827 GTGAAAACAGTTGTGAGCGCTCTGAACCCTGCTTCCCAATTTCGGCCCCT 876

276 GluGluAlaAspProAlaAlaGlyThrAlaCysLeuProPheTyrArgSe 292
    :::                    |||||||||||||:::|||||||||
877 CCAGGAGATCAGCGTCTGCGCCCTAACACCGTCTCCCTGTCTTCCGGTC 926

292 rSerArgAlaAlaCysGlyThrGlyAspGlnAspLeuPheGlyAsnLeuS 309
    |||:::                  |||||||||||||||||||:::|||
927 AGCACCGACTTGGTTCGGGCACACTGCCTACATGTTTGGTGAAGTC. 975

309 erThrArgAlaAsnProArgGlnGlnMetAsnGlyLeuThrSerPheLeuAsp 325
    |||   |||         :::|||    |||||||||||||||||||||||
976 .....CCAAATGTTCGGGAACAGGATCAATACTCTAACAGCTTACCTAGAT 1020

326 AlaSerThrValTyrGlySerSerProAlaLeuGluArgGlnLeuArgAs 342
    |||:::   |||||||||||||||||||||||   :::  :::|||||
1021 GCTGGACAAGTTTACGGTTCAGAGGATGGACTAGCAAAGAAACTGCGGGA 1070

342 nTrpThrSerAlaGluGlnMetAsnGlyValHisGlyArgLeuArgAspS 359
    |||||||||                   |||||||||||||||||||||
1071 CCTTACTAATGATGCTGGTCTTTTTACGTGTCAATAATCGTTTTAAAGACA 1120

359 erGlyArgAlaTyrLeuProPheValProProArgAlaProAlaAlaCys 375
```

```
831 uGlyAspAspGlyArgThrCysVal 839
    ||||:::||||::: ::||||||:::
2457 TGGTAGTGATGGTAAAACTTGCATG 2481
```

```
1121  :||::||||        |||
      ATGGACGAGAGCTCCTGCCCTTTACTAGTGTTAACACC...AACTTGTGC  1167

376   AlaProGluPro........GlyAsnProGlyGluThrArgGlyProCy  389
      |||            ::::::
1168  GCCACTCGTCAGAAAATCCTCAACGACAGCACTCTGACAGAGGTTCCTTG  1217

389   sPheLeuAlaGlyAspGlyArgAlaSerGluValProSerLeuThrAlaL  406
      ||||:||||||||:::||:::||||     |||
1218  CTTCATTGCGGGTGATGCTCGTGCAGTCGATGAGAATCCTGCTCGAACTCCT  1267

406   euHisThrLeuTrpLeuArgGluHisAsnArgLeuAlaAlaLeuLys  422
      ||||||||||||||::::::||||:::||| ||:::
1268  TACACACACTTTTGTAAGGGAGCACAATAGATTGGCTGCCCTGCCCTTCAT  1317

423   AlaLeuAsnAlaHisTrpSerAlaAspAlaValTyrGlnAspLeuAlaArgGly  439
      :::||| ::::::||||:: ||||||     ||||
1318  GTGCTCAACCCTACAATGGAGTGAAACTCTTTATCAAGAGCAAGAAA  1367

439   sValValGlyAlaLeuHisGlnIleIleThrLeuArgAspTyrIleProA  456
      |:::||||| ||||||:||||:::||| |||
1368  AATTGTTGGTGCATTCAACCAGATCTTGGTGATCAAGGAATAACTTGCCAC  1417

456   rgIleLeuGlyProGluAlaPheGlnGlnTyrValGlyProTyrGluGly  472
      ||:::||||| |||::||||:::||||||||
1418  TAATTGTAGGCCCTGATGTCTTACAATAGACACCTTGGACCATATCCAGGT  1467

473   TyrAspSerThrAlaAsnProThrValSerAsnValPheSerThrAlaAl  489
      |||||||| |||||||:::||||||
1468  TACAATGAGATGTGACCCTACCATTGCGACCGTCTTTGCCACTGCTGC  1517

489   aPheArgPheGlyHisAlaThrIleHisProLeuValArgArgLeuAspA  506
      ||||||||| |||:::||| |||:::
1518  TTTTCCGTTTTGCCCACTGACTATCCAGCCCTTCATTTTCGTCTTGATG  1567

506   laSerPheGlnGluHisProAspProGlyLeuTrpLeuHisGlnAsnAla  522
      :::::::||| |||:::|||||||||
1568  AAAACTACAAGAACCTCAGTTTCCCAGTGTCCCCTCTATGAGGCT  1617

523   PhePheSerProThrTrpLeuArgGlyGlyGlyGlyLeuAspProLeuI  539
      ||||||||||||     ::::::|||
1618  TTCTTCTCCACTGAGGGGTGATCTTGAAGGGGAATTGATCCTGTGCT  1667

539   eArgGlyLeuAlaArgProAlaLysLeuGlnValGlnAsnLeuAspLeuM  556
      |||||||:::|||||| ||||||
1668  CCGTGGATTGATTGGTCGTCCGGCAAAACTGAACACCCAGGACCACATGT  1717

556   etAsnGluLeuThrGluLeuGluPheLeuSerValLeuSerAsnSerSerThr  572
      ::                 ||||||||||||
1718  TGGGTAAGTCTCTCAGAGAGGCGTGTTTTGGCCTTCACATCCCACATAGCT  1767

573   LeuAspLeuAlaSerIleAsnLeuGlnAsnLeuGlyArgAspHisGlyLeuPr  589
      |||||||||||||||:::||:::|||||||
1768  TTGGATCTGGCTTCCCTCAACATGCAAAGAGGGCGTGACCATGCTATACC  1817

589   oGlyTyrAsnGluTrpArgGluPheCysGlyLeuProArgGluLeuThrP  606
      :::|||||||||:::::|||||:::|||
1818  AGGTTATAATGCATGGCGTCGGTTCTGTGGATGTTGTCGCTCCTCCAGAAATG  1867

606   roAlaAspSerThrAlaIleAlaSerArgSerValAlaAspLysIle  622
      :::|||::: :::||||:::
1868  AGCAAGAATTGGCTGTGGTGATGAACAACTAAACTAGCCCGCAAGTTA  1917

623   LeuAspLeuTyrLysHisProAspAsnIleAspValTrpLeuGlyGlyLe  639
      :::|||||| ::::::
1918  ATTGAGCTTTATGGACCACCCTGAGAACATTGATGTTTGGTTAGGAGGTGT  1967

639   uAlaGluAsnPheLeuProArgAlaArgThrGlyProLeuPheAlaCysL  656
      |||::::::||||||
1968  TGCTGAGCCGTTTTGCTCCTGCGGTTCTCTTTTTGCCTGCC  2017

656   euIleGlyLysMetLysAlaLeuAlaLeuArgAspGlyAspPheThrTrpTrp  672
      ||||::::::   :::|||:::
```

```
2018  TGATTCAAGACAGTTCCAGAAATCCGAGATGGCGATAGGTTGTGGTTT  2067
                                              |||
673   GluAsnSerHisValPheThrAspAlaGlnArgArgGluLeuGluLysHi  689
      |||::::: ||||||||| |||  :::     |||
2068  GAGACAATGGAGTGTTCACCAACCAAACAGAAAACAGCTTCGGCTTCTGT  2117

689   sSerLeuSerArgValIleCysAspAsnThrGlyLeuThrArgValProM  706
      ||||::|||||:||||:::::|||| ||||||
2118  GTCAATGGGCCGCATAATCTGCGACAACACGGAATCCTAAAGTTCCCA  2167

706   etAspAlaPheGlnValGlyLysPhePheGluProGluAspSerCysAsp  722
      ||| ||| :|||||| |||||
2168  GAGATCCTTTCGCTTC...CGGAGTCCTGCCAGTTTTGTTAACTGTGGG  2214

723   SerIleThrGlyMetAsnLeuGluAlaTrpArgGluThr..........  735
      ::||| ::::::|||||| |||:::||||||
2215  AATATCCCAGCTTTTGATCTCGAGCCTTGGAAAGAGACTGGTGACGGTGG  2264

736   ...................PheProGlnAspAspL  741
      ::|||::::::|||
2265  CATCAACATTCAGACATTGGAAATGAGATAGAGATTCCATCCCAGGACC  2314

741   ysCysGlyPheProGluSerValGluAsnGlyAspPheValHisCysGlu  757
      :::                     |||||||
2315  TCCCGGACCCAGGGGACCTCCAGGAGAAC....GTGGTCCACAGGGTG  2358

758   GluSerGlyArgArgValLeuValTyrSerCysArgHisGlyTyrGluLe  774
      ::|||                          GACCT
2359  TAGCAGGGCCACCAG...............GACCT  2378

774   uGlnGlyArgGluGlnLeuThrCysThrGlnGluGlyTrpAspPheGlnP  791
      |||| ::::::||| |||:::
2379  CCAGGAATTCCAGGACCTCCCATTAATACACAG.....GACAACAGTC  2422

791   roPro.LeuCysLysAspValAsnGluCysValAlaAspGlyAlaHisPr  807
      :::||||
2423  TGCCTTCTT................TGCCTCAGTCAACTCACTCCTTC  2454

807   oCysHis.AlaSerAla.  812
      ||||| |||:::
2455  CTGCCACTGCTAAGTTGGTGGTTGTCGGTCCGGTCCTTTACAAATGGACAA  2504

813   ..ArgCysArgAsnThrLysGlyPheGlnCysLeuCysAlaAspPr  828
      :::     |||:::||| |||
2505  AACCATTATAACCAAACCTCAGGGATGTTCTTGTGCCAGATACCTGGTGT  2554

828   oTyrGluLeuGlyAspAspGlyArgThrCysValAspSerGlyArgLeuP  845
      ||||||     :::|||
2555  ATATGAATTC.........GAGTTTTCATCCATAGGGACTCGTAGTTTGG  2595

845   roArgValThr  848
      |||||
2596  GCTTTGTGACC  2606
```

seq_name: gb_ov:AF378824

```
JOURNAL     Blood (2001) In press
REFERENCE   2 (bases 1 to 2814)
AUTHORS     Crowhurst,M.O., Lieschke,G.J., Layton,J.E. and Ward,A.C.
TITLE       Direct Submission
JOURNAL     Submitted (10-MAY-2001) Cytokine Biology Laboratory, Ludwig
            Institute for Cancer Research, The Royal Melbourne Hospital,
            Parkville, Vic 3050, Australia
FEATURES             Location/Qualifiers
     source          1..2814
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /chromosome="LG X"
                     /map="between z8146 and z9473"
                     /tissue_type="kidney"
                     /dev_stage="adult"
                     /note="sequence AF326958 deposited in GenBank was used as
                     a probe"
     gene            <1..2814
                     /gene="mpx"
     CDS             <1..2037
                     /gene="mpx"
                     /note="myeloperoxidase; similar to mammalian eosinophil
                     peroxidase; alternatively spliced"
                     /codon_start=1
                     /product="myeloid-specific peroxidase"
                     /protein_id="AAK91661.1"
                     /db_xref="GI:15193021"
                     /translation="DYMAQTLRLISEKAHHVHKRSINATNLLTQDELQTITRMTGCEA
                     QTRPPSCRTRPLINKYRTASGVCNNRKNPLLGASNTPFARWLPAQYEDAVSQPKGWDP
                     NKLYNGAALPMVRLVSNRILATADADIESDHDFTFMLTIFGQWVDHDLTFTPFSPSIR
                     SFSNGIDCENSCERSEPCFPISAPPGDQRLRPNTCLPVFRSAPTCGSGHTAYMFGEVP
                     NVREQINTLTAYLDAQGVYGSEDGLAKELRDLTNDGGLLRVNNRFKDNGRELLPFTSV
                     NTNLCATRQKILNDSLTEVPCFIAGDARVDENPALNSLHTLFVREHNRLARALHVLN
                     PTWSSEFLYQEARKIVGAFNQIVIKEYLPLIVGPDAYNRHLGPYGNENVDPTIAN
                     VFATAAFRFAHLTIQPFIFRLDENYKNHPQFPSVPLYEAFFSPWRVIPEGGIDPVLRG
                     LIGRPAKLNTODHMLVNALRERLFAFTSHIALDLASLNMQRGRDHAIPGYNAWRRFCG
                     LSAPQNEQELAVVMNNTELARKLIELYGTPENIDVWLGGVAEPFAPGGRVGSLFACLI
                     SRQFQKIRDGDRLWFESNGVFTTKQKTALASVSMARIICDNTGILKVPRDPFRFRSPA
                     SFVNCGNIPAFDLEPWKETGDGGINISDIGNEIEIPSQDLPDPGDLQENVVHRV"
BASE COUNT      701 a    658 c    618 g    837 t
ORIGIN
```

alignment_scores:
```
        Quality:  1619.00              Length:       742
        Ratio:       3.009             Gaps:          10
Percent Similarity: 72.507      Percent Identity: 42.588
```

alignment_block:
```
US-08-482-402A-3_COPY_1_848  x  AF378824

Align seg 1/1  to: AF378824    from: 1  to: 2814

   92 GluIleMetGluThrSerIleGlnAlaMetLysArgLysValAsnLeuLeu 108
      :  |||      ::::||     |:    ::::::||||:::::
    1 GACTACATGGCACAAACGCTGAGACTGAGACTCAGTGAGAAAGCCCATCATGT 50

  108 sThrGlnSerGlnHisProThrAspAlaLeuSerGluAspLeuLeuS 125
      :::::||||   ::  |||:::|  ||:::::||||::::::||  |||
   51 ACACAGAGAGTCCATCAATGCCACAAACCTGCTCACTCAAGACGAGCTTC 100

  125 erIleIleAlaAsnMetSerGlyCysLeuProTyrMetLeuProProLys 141
      ||:  ||:|:::::|||::|||||       |||||||:::
  101 AAACCATTACACGTATGACAGGCTGTGAGCTCAAACCCGGCCACCATCT 150

  142 CysProAsnThrCysLeuAlaAsnLysTyrArgProIleThrGlyAlaCy 158
      :::   |||    |||   ||:||:||:  ||||||||||::|||::
  151 TGTAGGACCAGACCCCTCATAAACAAGTATCGAACTGCCAGCGGTGTCTG 200

  158 sAsnAsnArgAspHisProArgTrpGlyAlaSerAsnThrAlaLeuAlaA 175
      |||||||||   :::||||  ||||||||||||||      :::|||
  201 CAACAATCGGAAGAACCCACCTTCTTGGTGCATCCAACACTCCATTTGCCC 250


  175 rgTrpLeuProProValTyrGluAspGlyPheSerGlnProArgGlyTrp 191
      ||||||||||     |||||||||:::  |||||||  :::  :::||||
  251 GCTGGTTGCCTGCTCAATATGAGGACGGCGCCGTTCTCAGCCAAAAGGATGG 300

  192 AsnProGlyPheLeuTyrAsnGlyMetAlaTrpProLeuProProValArgGluVa 208
      :::||:   ||||||||:::::  :|||  :::|||||||::::||::
  301 GATCCCAACAAGTTGTACAATGGTGCAGCGCTGCCCATGGTTAGACTGGT 350

  208 lThrArgHisValIleGlnValSerArgAsnGluValValThrAspAspAspA 225
      |:::::::::||||    ||  |:::|| ::::  :::|||
  351 TTCCAATCGCATTCAGCCACTGCAGATGCTGACATAGAGAGTGACCATG 400

  225 rgTyrSerAspLeuLeuMetAlaTrpGlyGlnTyrIleGluAspHisAspIle 241
      :::::||       |||:::|||||||||||:::::  |:::||||||||:
  401 ACTTTACCTTCATGCTCACCCATCTTTGGGCAATGGTTGACCACGATCTG 450

  242 AlaPheThrProGlnSerThrSerLysAlaAlaPheGlyGlyLeuSerArgS 258
      :::::||     |||:::|||   ::||    |:::  |||||||:::|||
  451 ACTTTACACCTTTCTCACCAAGTATTAGGTCCTTTAGTAATGGCATTGA 500

  258 pCysGlnMetThrCysGluAsnGlnAsnProCysPheProIleGlnLeuP 275
      ||||:::    ||   ::||||   :::|:::||||:::::||  :::
  501 CTGTGAAAACAGTTGTGAGCGCTGCAACCCTGCTTCCCAATTTCGGCCC 550

  275 roGluGluAlaArgProAlaAlaGlyThrAlaCysLeuProPheTyrArg 291
      ||   ::::::::|:::||||||:::||||||:::||||||:::::||:  :::||
  551 CTCCAGGAGGATCAGCGTCTGCGCCCTAACACCTGCTCCCGTCTTCCGG 600

  292 SerSerAlaAlaCysGlyThrGlyAspGlnGlyAlaLeuPheGlyAsnLe 308
      :::::::::||:||::||||::::||  :::|||  :::|||  :::|||
  601 TCAGCACCGACTTGTGGTTCTGGGCACACTGCCTACATGTTTGGTGAAGT 650

  308 uSerThrAlaAsnProArgGlnGlnMetAsnGlyLeuThrSerPheLeuA 325
      ||:::::||||    :::||  ::::::|||||||::||||||||:::|||
  651 C......CCAAATGTTCGGGACAGATCAATACTCTAACAGCTTACCTAG 694

  325 spAlaSerThrValTyrGlySerSerProAlaLeuGluArgGlnLeuArg 341
      ||||:::    :::::::|||||||||||  |||    |||:::|||:::
  695 ATGCTGGACAAGTTACGGTTCAGGAGGATGGACTAGCAAAGGAACTGCGG 744

  342 AsnTrpThrSerAlaGluGlyLeuLeuArgValHisGlyArgLeuArgAs 358
      :::::||    :::||||   :::||    ::||||||  :::  ::::::
  745 GACTTACTAATGATGGTGGCCTTTACGTGTTAATAATCGTTTAAAGA 794

  358 pSerGlyArgAlaTyrLeuProPheValProProArgAlaProAlaAlaC 375
      |:::|||   |:::   ||||||||:::||   |||:::||    :::
  795 CAATGGACGAGAGCTCCTGCCCTTTACTAGTGTTAACACC...AACTTGT 841

  375 ysAlaProGluPro........GlyAsnProGlyGluThrArgGlyPro 388
      ||||| :::        :::::  |||  |||  :::||||||:::||
  842 GCGCCACTCGTCAGAAAATCCTCAACGACAGCACTCTCACAGAGGTTCCT 891

  389 CysPheLeuAlaGlyAlaAspGlyArgAlaSerGluValProSerLeuThrArgL 405
      ||||||:::::||||   :::||||||  :::|||:::|||:::  :::::
  892 TGCTTTATTGCAGGTGATCCTCGTGTCGATGAATCCTGTCCTGAACTC 941

  405 aLeuHisThrLeuTrpLeuArgGluHisAsnArgLeuAlaLeuAlaLeuL 422
      :::|||    |||||||:::|||||||||||||||||||||   :::
  942 CTTACACACTTTTGTAAGGGAGCAAATAGATTGGCTCGCGCCTTC 991

  422 ysAlaLeuAsnAlaHisTrpSerAlaAspAlaValTyrGlnGluAlaArg 438
      |:::||    :::||  :::|||  |:::||||||||||||||||
  992 ATGTGCTCAACCCTACCTGGAGTAGTGAAACTCTTTATCAAGAGCAAGA 1041

  439 LysValValGlyAlaLeuHisGlnIleIleThrLeuArgAspTyrIlePr 455
      :::::||||||||:::|||   :::|||:::   :::||||||:::::
  1042 AAAATTGTCGGTGCATTCAACCAGATCTTGGTGATCAAGGAATACTTGCC 1091

  455 oArgIleLeuGlyProGluAlaPheGlnTyrValGlyProTyrGluG 472
      :::|||||||||||||    |||||  :::   |||:::||||||
  1092 ACTAATTGTAGGCCCTGATGCTTACAATAGACACCTTGGACCATATCCAG 1141

  472 lyTyrAspSerThrAlaAsnProThrValSerAsnValPheSerThrArgAla 488
```

```
2033 TGTAGCAGGGCCACCAG.................................GAC 2052

774 LeuGlnGlyArgGluGlnLeuThrCysThrGlnGluGlyTrpAspPheGl 790
     ||||||     |||  ::::|||            |||   ::
2053 CTCCAGGAATTCCAGGACCTCCCATTAATACAACAG.....GACAACAG 2096

790 nProPro.LeuCysLysAspValAsnGluCysValaAspGlyAlaHisPro 806
     :  ||| |||                       ||| :::  |||||
2097 TCTGCCTTCTT...............TGCCTCAGTCAACTCCATCCT 2128

807 ProCysHisAlaSerAlaArgCys 814
     ||||||     ::: |||
2129 TCCTGCCACTGCTAAAGTTGTTGT 2152
```

```
1142 |||||:::::::::|||||:::|||||::::::|||||||||::|||||||
     GTTACAATGAGAAATGTAGACCCTACCATTGCGAACGTCTTTGCCACTGCT 1191

489 AlaPheArgPheGlyHisAlaThrIleHisProLeuValArgArgLeuAs 505
    ||||||||||||||  :||| |||     ||||||||::||:::|::
1192 GCTTTCCGTTTCGGCCACTGACTATCCAGCCCTTCATTTTCGTCTTGA 1241

505 pAlaSerPheGlnGluHisProAspLeuProGlyLeuTrpLeuHisGlnA 522
    |  |||||:||||| ||||||||::|||||||| :|||||||||
1242 TGAAACTACAAGAACCATCCTCAGTTTCCCAGTGTCCCCCTCTATGAGG 1291

522 laPhePheSerProTrpThrLeuLeuArgGlyGlyGlyLeuAspProLeu 538
    ||||||||||||||   :::::   ||||||:::||||:::
1292 CTTTCTTCTCTCCATGGAGGTGATCTTTGAAGGGGAATTGATCCTGTG 1341

539 IleArgGlyLeuLeuAlaArgProAlaLysLeuGlnValGlnAspGlnLe 555
    :::|||||||||||||   ||| |||:::||||||||||:::|||||
1342 CTCCCTGGATTGATTGGTCGTCCGGCAAAACTGAACACCCAGGACCACAT 1391

555 uMetAsnGluLeuThrGluArgLeuPheLeuSerAsnSerSerT 572
    :::    ||| :::|  |||||||||||||||:::|||||:::
1392 GTTGGTGAATGCTCTCAGAGAGGCTGTTTGCCTTCACATCCCACATAG 1441

572 hrLeuAspLeuAlaSerIleAsnGlnAsnLeuArgGlyHisAspLeuGlyLeu 588
    :::||||||||||:::|||  :::||:::|||||||||||||:::::::
1442 CTTTGGATCTGGCTTCCCTCAACATGCAAAGAGGGCCGTGACCATGCTATA 1491

589 ProGlyTyrAsnGluTrpArgGluCysGlyLeuProArgGlyLeuGluTh 605
    |||||||:|||||||||::|||  ||||:::
1492 CCAGTTATAATGCATGGCGTCGTGGATTGTCTGCTCCTCAGAA 1541

605 rProAlaAspLeuSerThrAlaIleAlaSerArgSerValAlaAspLysI 622
    :  ||||||:::::|||||:::   ||| :::|||
1542 TGAGCAAGAATTGGCTGTGGTGATGAACAACACTGAACTAGCCCGCAAGT 1591

622 leLeuAspLeuTyrLysHisProAspAsnIleAspValTrpLeuGlyGly 638
    ::::::|||||||    ||||||||||||||||||||
1592 TAATTGAGCTTTATGGCACCCCTGAGAACATTGATGTTTTCGTTAGGAGGT 1641

639 LeuAlaGluAsnPheThrPheAspAlaGlnThrGlyArgGlyProHeAlaCy 655
    ::||||||   ::|||  |||:::|||    |||||||||
1642 GTTGCTGAGCCTTTGCCTCCTGGTCGGTCGGTTCTTCTTTTTGCCTG 1691

655 sLeuIleGlyLysGlnMetLysAlaLeuAspGlyAspTrpPheTrpT 672
    ::|||||||::||  ||||||||||||||||||  ::||:
1692 CCTGATTTCAAGACAGTTCCAGAAAATCCGAGATGGCCATAGGTTGTGGT 1741

672 rpGluAsnSerHisValPheThrAspAlaGlnAsnArgArgGluGluLys 688
    :::||||   ||||||||||||   |||  :::||
1742 TTGAGAGCAATGGAGTGTCACCCAACAGAAAAACAGCTCTGGCTTCT 1791

689 HisSerLeuSerArgValIleCysAspAsnThrGlyLeuThrArgValPr 705
    ||  ::::::|||||||:::|||||||||||||:::  :::|||
1792 GTGTCAATGGCCCCGCATAATCTGCCACACCGGAATCCTTAAAGTTCC 1841

705 oMetAspAlaPheGlnValGlyLysPheProGluAspPheGluSerCysA 722
    :::|||||||||||::|||   :::
1842 CAGAGATCCTTTCGCTTC...CGGAGTCCTGCCAGTTTGTTAACTGTG 1888

722 spSerIleThrGlyMetAsnLeuGluAlaTrpArgGluThr......... 735
    :|||   ||  :::|||||||||    :::|
1889 GAAATATCCCAGCTTTGATCTCGAGCCTTGGAAAGAGACTGGTGACGGT 1938

736 ..................PheProGlnAspAs 740
                      :::|||::::::||
1939 GGCATCAACATTTCAGACAATTGGAAATGAGATAGAGATTCCATCCCAGGA 1988

740 pLysCysGlyPheProGluSerValGluAsnGlyValHisCysG 757
    ||||||                 :::||||||
1989 CCTCCCGACCCAGGGGACCCTCCAGGAGAAC......GTGGTCCACAGGG 2032

757 luGluSerGlyArgArgValLeuValTyrSerCysArgHisGlyTyrGlu 773
    :::||:::::::
```